A Location-Aware Nutritional Needs Prediction Tool for Type II Diabetic Patients: Case Kenya

By

Lulu Amina Karega

124272

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information Technology at Strathmore University

School of Computing and Engineering Sciences
Strathmore University
Nairobi, Kenya

October, 2022

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Lulu Amina Karega

Sign:  Date: 23rd June 2022

Approval

The thesis of Lulu Amina Karega was reviewed and approved for examination by the following:

Dr. Vincent Omwenga
Research Director, Senior Lecturer
School of Computing and Engineering Sciences
Strathmore University

Dr. Julius Bitume
Dean, School of Computing and Engineering Sciences
Strathmore University

Dr. Bernard Shibwabo
Director of Graduate Studies
Strathmore University

Abstract

Diabetes is a chronic disease caused by a lack of insulin production by the pancreas or by poor utilization of the insulin that is produced, with insulin being the hormone that helps glucose get to blood cells and produce energy. Urbanization and busy day to day schedules mean patients tend to pay little or no attention to their dietary habits which results in a preference for fast foods and processed food. The prevalence of type II diabetes in the world, Kenya included, has been steadily rising over the years and is projected to keep growing at an alarming rate. Diabetes if not properly managed can result in long-standing, costly and time-consuming complications. Diabetes management and control of blood sugar levels are generally done by the use of medication, namely insulin and oral hypoglycaemic agents. However nutritional therapy can also go a long way to boosting the general health of a patient and reducing risk factors leading to further complications. Personalised nutrition has been formally defined as healthy eating advice, tailored to suit an individual based on genetic data, and alternatively on personal health status, lifestyle, and nutrient intake. Diabetes management falls under the field of health informatics that can benefit from data analytics. Predictive analytics is the process of utilizing statistical algorithms, software tools and services to analyze, interpret and visualize data with the aim to forecast trends, and predict data patterns and behaviour within or outside the observed data. This study sought to develop a location-aware nutritional needs prediction tool for type II diabetic patients in Kenya. The prediction tool would help both nutritionists and patients by providing accurate and relevant nutritional advice that would help in dietary changes to combat type II diabetes with the added benefit of being location aware. The tool will use pathological results from nutritional testing to support nutritional therapy. If any deficiencies are identified from the provided nutritional markers, food items likely to improve those nutrient levels will be recommended. The amount of nutrient available in a given food item are determined by the food composition table for Kenya as published by the Food and Agriculture Organization (FAO) in conjunction with the Kenyan government. The study used a simplistic implementation of matrix factorization to provide predictions of locally available food items, down to the county level.

Keywords*: Health Informatics, Nutritional Therapy, Predictive Analytics, Diabetes*

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
CSS	Cascading Style Sheets
FAO	Food and Agriculture Organization
FCT	Food Composition Table
GBM	Gradient Boosted Model
GLM	Generalized Linear Model
GOK	Government of Kenya
GPU	Graphics Processing Unit
HTML	HyperText Markup Language
IDF	International Diabetes Federation
KNBS	Kenya National Bureau of Statistics
ML	Machine Learning
MNT	Medical Nutritional Therapy
NCD	Non-Communicable Diseases
PaaS	Platform as a Service
SaaS	Software as a Service
SDLC	Software Development Lifecycle
SVM	Support Vector Machines
UML	Unified Modelling Language
WHO	World Health Organization

Acknowledgements

I would like to acknowledge the Almighty God for the strength and grace to pursue this master's degree, particularly during these unprecedented times. My sincere gratitude to the members of the School of Computing and Engineering Sciences staff, particularly my thesis supervisor, Dr Vincent Omwenga for his continued guidance and support from the inception to the completion of this study.

Dedication

This thesis is dedicated to my family for their never-ending support and encouragement during this research. My father, Dr R. M. Karega and my late mother, Ms Patience Ponda who not only raised and nurtured me but also taxed themselves dearly over the years for my education and intellectual development. My brother, Mohamed Karega, who never lost faith in me through this journey. My classmates, who have been a source of motivation and strength during moments of despair and with whom we collectively managed to help each other power through. Finally, I acknowledge my husband, Badi Weru, for his love and tremendous support.

Chapter 1: Introduction

1.1 Background of the Study

Non-communicable diseases (NCDs) are the leading global causes of death, accounting for almost 70% of all deaths worldwide. Data shows that of these deaths, 80% occur in low-income and middle-income countries (World Health Organization, 2010). An approximated 422 million people worldwide have diabetes mellitus, and 1.6 million deaths annually are as a result of diabetes making it the 4th leading contributor of NCD-related deaths (Damasceno, 2016).

In Kenya, while Communicable Diseases still account for the majority of the disease and mortality burden in the country, the impact and prevalence of NCDs, diabetes in particular, has been rising. For those living with diabetes, access to affordable treatment is crucial to survival. Dr Nancy Ngugi, Head of the Diabetics Clinic at Kenyatta National Hospital noted, "Across Kenya, community awareness around Diabetes is low." (WHO, 2014). In most cases, the diagnosis is made when the patients visit the hospital due to complications resulting from diabetes such as thirst, fatigue, and constant hunger or through medical outreach camps. Unfortunately, for most Kenyans, the cost and access to healthcare are a hindrance to treatment. As Dr Ngugi said, "Healthcare is not free in Kenya. In low-income communities, we find that life is hard. People do not have money to buy drugs, to see the doctor, to get their tests done." (WHO, 2014). It therefore would be helpful to assess the increasing burden of diabetes with the aim to provide cost-effective strategies for its prevention and management.

The prevalence of diabetes in adults is 3.3% and this is expected to rise to 4.5% by 2025 (O'Hara et al., 2016) However, this could be on the lower side as up to two-thirds of diabetics are possibly undiagnosed (Jones, 2013). These alarming statistics led the Kenyan Government to institute measures such as the establishment of bodies mandated to spearhead the campaign on diabetes management. Concern has been raised over the high prevalence of pre-diabetes and undiagnosed cases which can easily spiral into complications such as cardiovascular damage, kidney and nerve damage, poor eyesight, or even death (Mohammed et al., 2018; English, 2021).

Commented [M1]: Formatting

The Diabetes Management Act of 2013, tabled by the National Diabetes Council, states in Article (I) section 3 (a) that one of its purposes is to "promote public awareness about the causes, consequences, means of prevention, treatment, and control of diabetes in Kenya". Underlying these measures is the credence that an empowered citizenry will be an effective tool

to combat the challenges associated with diabetes management making nutrition education one of the most effective strategies in diabetes management and prevention.

Focusing on nutritional management of type II diabetes in Kenya, a key challenge has been getting information on nutritional requirements and the sources available within a patient's locality. According to the Kenya Nutritionists and Dieticians Institute, there are 2167 registered nutritionists and dieticians. Given the population of 47.5million (Kenya National Bureau of Statistics, 2019a) this is a service ratio of approximately 1:20000. Access to personalized nutritional information on the management of type II diabetes would go a long way in helping with nutritional therapy by providing information that is not only specific to an individual's needs but also tailored to resources available in their locality or within reasonable reach.

This study looked at existing predictive analysis tools and techniques, especially those geared toward health and nutrition to determine how to best leverage them to provide this information given a patient's nutrition pathological profile. Predictive analytics is defined as the process of utilizing statistical algorithms, software tools and services to analyze, interpret and visualize data with the aim to forecast trends, predict data patterns and behaviour within or outside the observed data (Dinov, 2018)

1.2 Problem Statement

There were 8,700 deaths due to diabetes in Kenya in 2015 (Shannon et al., 2019). This figure is probably an underestimation, Dr. Roglic, who leads the World Health Organization's work on diabetes says, "Most people with diabetes do not die of causes uniquely related to diabetes, but of associated cardiovascular complications, like a heart attack" (WHO, 2014). Most of these suffer from type II diabetes which is often preventable and results from the body not utilizing insulin effectively.

Organizations like Novo Nordisk (Shannon et al., 2019) have in the past supported projects and initiatives aimed at providing community-based management and awareness of diabetes (Shannon et al., 2019). However, their reach has been limited geographically due to a lack of conclusive knowledge of where to target these initiatives.

The main challenge for the average Kenyan is accessing the relevant information and affordable healthcare for people living with diabetes. This is problematic on two fronts; one is the mentioned scarcity of dieticians and nutritionists. The second challenge is availing information of the same standard and quality across the board, with respect to the vastly varying socio-economic statuses. As per the last census report, only 4.3% of the population aged of 15 and

Commented [M2]: typos

above have searched for goods or services online; this number dips to 1.7% when the focus is on rural areas (Kenya National Bureau of Statistics, 2019b).

This means even with patients having access to their nutritional profiles, it is difficult for them to get accurate and reliable nutritional information for identifiable deficiencies, such as what food to eat to address given deficiencies. It is also a challenge for nutritionists who may not be aware of the macronutrient composition of food items within a specific location and surrounding areas. This information would help nutritionists give accurate and relevant nutritional advice to patients that would help in dietary changes to combat type II diabetes and its resulting complications.

1.3. Objectives

1.3.1 General objectives

This study aimed to develop a location-aware nutritional needs prediction tool for type II diabetic patients, to help patients get access to accurate and reliable information based on nutritional deficiency needs.

1.3.2 Specific objectives

i. To investigate the challenges of providing diabetes nutritional information to the public.
ii. To evaluate existing techniques for predictions of suitable food items based on nutritional markers.
iii. To develop a nutritional needs prediction tool based on location and nutritional deficiency needs.
iv. To test the ability of the tool to give locale-specific predictions given nutritional markers.

Commented [M3]: predictions of what?

1.4. Research Questions

i. How does the public currently access nutritional information about diabetes?
ii. What existing nutritional prediction tools systems are there?
iii. What algorithm or technique would be best suited to build a location-aware nutritional needs prediction tool based on nutritional deficiency needs?
iv. How will the prediction tool be validated?

1.5. Justification

This study is useful as it will enable the provision of information for the nutritional management of diabetes, thereby increasing public awareness of the disease, its prevention and management.

This will in turn decrease the number of prediabetes and undiagnosed diabetes cases since a healthy lifestyle and proper dietary intake greatly reduce risk factors for diabetes It would also lead to decreased cases of disease-related complications and hopefully lead to an increase in preventive measures and relevant lifestyle changes being applied by those who already have type II diabetes. The study will help researchers and students by adding to the knowledge base in this area of health informatics. The developed tool can be leveraged by other researchers in other locations as this tool will be tailored for Kenya.

1.6. Scope and Limitation

The focus of this study was on constructing a nutritional needs prediction tool that will respond with relevant dietary predictions data given a location and certain nutritional markers.

The system was made accessible via several devices to improve the means of accessibility. A web-based application was used as they are versatile and can be accessed across a range of devices including portable ones, such as mobile phones and tablets.

This study did not come up with any nutritional information, merely collate existing information and validate it with the assistance of nutritionist(s) registered with the Kenya Nutritionists and Dieticians Institute.

The study was geographically limited to Kenya, with the results from the search engine tailored to offer optimal information for the given location which has to be within Kenya.

Chapter 2: Literature Review

2.1. Introduction

With the advancement of information technologies, there have been several new fields emerging such as artificial intelligence (AI) and big data analysis. Analytics is the process of using computational methods to discover and report patterns in data so as to gain insight and affect decisions (Abbott, 2014). Analytics falls into four categories:

- Descriptive analytics: Analyses historical raw data to describe what happened or what is happening
- Diagnostic analytics: Looks at data with the intention to uncover correlations between variables and recognizing causal relationships
- Predictive analytics: This is the basis of the study and refers to the process of discovering meaningful patterns in historical or existing data and using them to make forecast future or unknown events
- Prescriptive analytics: Look at the various factors in a scenario and suggests takeaways, it tends to be a combination of data and business rules

Commented [M4]: This does not describe predictive analytics

2.2 Predictive Analytics

Commented [M5]: Why did you adopt predictive analytics?

Predictive analytics is a subset of analytics defined as the process of utilizing statistical algorithms, software tools and services to analyze, interpret and visualize data with the aim to forecast trends, and predict data patterns and behaviour within or outside the observed data (Dinov, 2018). Predictive analytics tools use data to identify the probabilities of possible outcomes. It was well suited to the study because predictive systems are used to augment human decision making, essentially by suggesting relevant items for a user to choose from a plethora of options(Hsu, 2014; Prabhu et al., 2019). In this case, the food composition tables offer a vast selection of food items the user or patient could select from but the predictive tools results in a much smaller ranked list in order of nutritional composition that the user can select from.

In this section, the different techniques and algorithms employed by predictive tools and their use in personalised nutrition are reviewed. Predictive analytics tools are powered by different models and algorithms which are reviewed below.

2.2.1 Predictive Analytics Models

Predictive analytics models are either parametric or non-parametric. Parametric models are based on the assumption that there are known distributions in the data and find linear relationships in the data. Non-parametric models do not have any underlying assumptions about the distribution of the data. Non-parametric models are more flexible but cannot guarantee optimal solutions while parametric models can be proven to have certainty as extensive data properties are known. The downside of parametric models is that they are time-intensive when it comes to transforming the data (Abbott, 2014; Babcock, 2016).

2.2.1.1 Classification Model

Classification models group data based on patterns in historical data. Very simplistic and easy to retrain making them suitable for different industries. Best used in guiding decisive actions such as determining whether a client will default on a loan, or if a customer should be churned from the system.

2.2.1.2 Clustering Model

Clustering models sort data into separate nested groups based on similar attributes. Used in retail to create targeted marketing strategies, ideally by separating customers into smaller groups based on commonalities.

2.2.1.3 Forecast Model

The forecast model works on metric value prediction by estimating the numeric value for new data based on patterns discovered in historical data. It also considers multiple input parameters and is applied in various fields e.g., predicting the volume of traffic to a website or the number of support calls to a call centre.

2.2.1.4 Outliers Model

Outlier models identify anomalous data entries in a data set. The anomalies can either be individual or together with other entries or categories. Often used in retail and finance to identify fraudulent transactions.

2.2.1.5 Time Series Model

Time series models comprise a sequence of data points with time as an input parameter. It is powerful in determining the evolution of a single metric over time with an increased level of

accuracy. It can be used to project a company's earnings over time by processing sales from previous time periods.

2.2.2 Machine Learning Algorithms in Predictive Tools

Machine Learning is a technique where computers are programmed to optimize performance criteria from training or example data. It is applied in cases where a program cannot be coded to solve a certain problem without training data or experience (Alpaydin, 2010). There are four possible approaches (Knox, 2018) to ML depending on the data available

- Supervised Learning when there is labelled data, meaning there are outputs that can be ascertained to be correct values for the inputs. Learning is achieved by training a dataset similar to the input data (Pandey et al., 2019).
- Semi-Supervised Learning where most of the data is not labelled. This is representative of a large amount of real-world data as it can be costly to label every single data point. Semi-supervised learning provides a handy middle ground (Ravi et al., 2017).
- Unsupervised Learning where the data is wholly unlabeled. There is no visibility of any patterns or features within the data and it is up to the algorithm to identify what it can. Clustering algorithms are some of the methods that can be used when dealing with an unsupervised learning problem (Chollet, 2017).
- Reinforced Learning works in an interactive environment allowing the agent to learn through trial and error via feedback from its own experiences (Chollet, 2017). It can either be positive reinforcement which relies on a reward-based approach, or negative reinforcement which is a punishment-based approach.

In the recent past, a lot of interest has been generated in subfields related to AI and subfields in health informatics like public health, bioinformatics, medical imaging, etc.

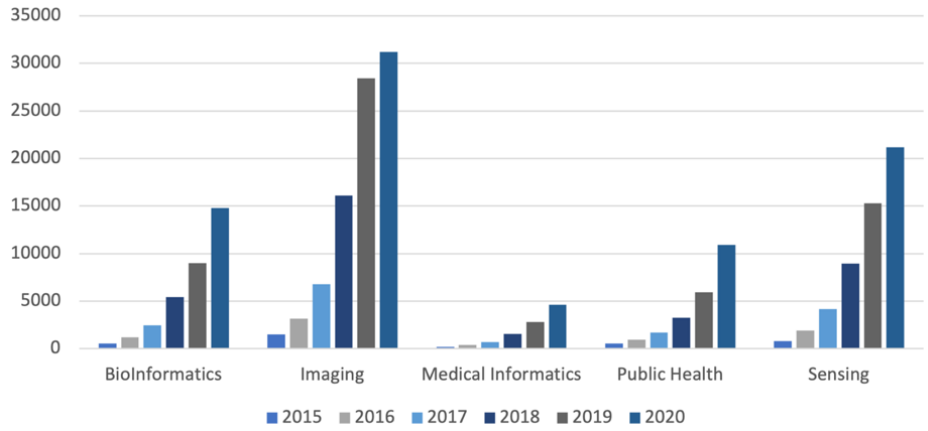


Figure 2.1: Distribution of papers published on machine learning in health informatics

Note. The statistics shown in Figure 2.6 were obtained from Google Scholar with the keywords: *"public learning"*, *"machine learning"*, *medical* *OR* *health* along with the annual time bounds.

2.2.2.1 Random Forest

The random forest algorithm is a combination of decision trees. Decision trees work by aggregating data into a tree with a root node that has sub-nodes and decision nodes that determine which leaf node to route to. Essentially a decision node is a question and a leaf node is an answer or consequence (Gorakala, 2016). With random forest, each tree depends on the values of a random vector sampled independently with the same distribution for all trees in the collective (*Predictive Analytics Models and Algorithms*, 2022).

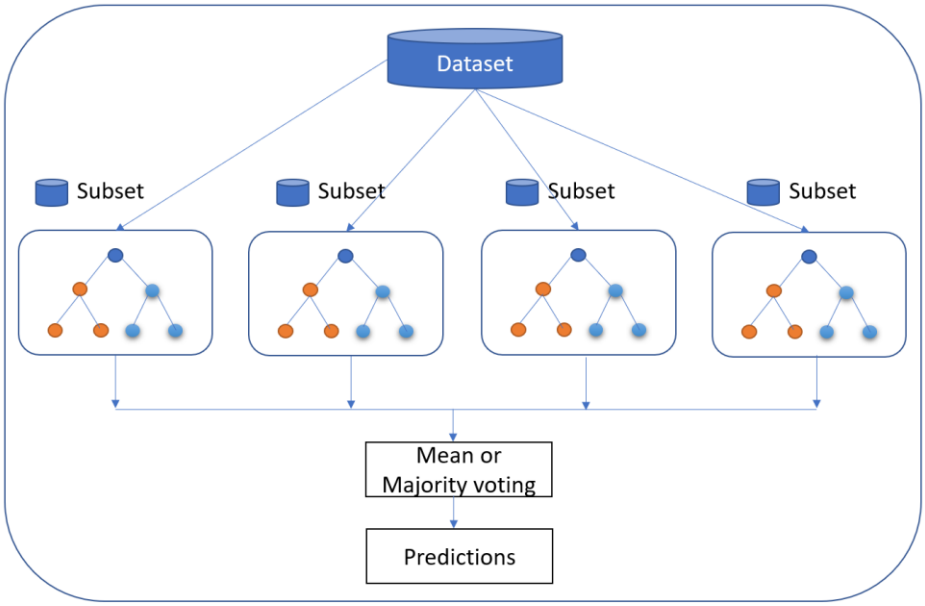


Figure 2.2: Illustration of Random Forest Algorithm

Note: From Predictive Analytics Models and Algorithms

2.2.2.2 Generalized Linear Model

GLM is a more complicated variation of the General Linear Model which takes the General Linear Model's analysis of the effects of multiple variables and then analyses different distributions so as to find a best fit model (Abbott, 2014; *Predictive Analytics Models and Algorithms*, 2022). A sample case would be analyzing the purchase of cooling fans given a rise in temperature, the general model could suggest that fifty fans are sold for every 5 degrees increase and therefore if the temperature went from 25 degrees to 35 degrees, a hundred fans would be sold. However, it would not be logical to conclude that if the temperature went up to 50 degrees, 250 units would be sold. GLM could suggest that sales might flatten once a certain temperature is reached. GLM has the advantage of training very quickly, however, it requires large datasets and is susceptible to outliers.

2.2.2.3 Gradient Boosted Model (GBM)

The Gradient Boosted Model is a prediction model that comprises of several decision trees, similar to the random forest algorithm. It uses the boosting machine learning technique to strengthen individual weak learners (decision trees) in the forest. The main difference between GBM and random forest is that it builds one tree at a time. The current tree helps correct errors in the previous tree unlike in random forest where there is no relation between individual trees. The main shortcoming of GBM is its speed, it is a slower performing algorithm because of building each tree individually (Babcock, 2016; *Predictive Analytics Models and Algorithms*, 2022).

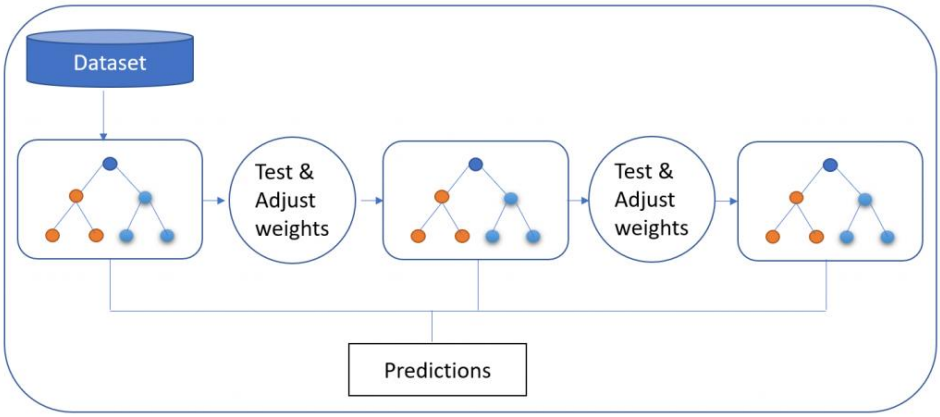


Figure 2.3: Illustration of Gradient Boosted Model Algorithm

Note: From Predictive Analytics Models and Algorithms

2.2.2.3 K-Means

K-Means is a high-speed algorithm that works by grouping unlabeled data points based on their similarities. It is used in clustering models and figures out common characteristics for individuals and groups them. It is particularly handy when dealing with a large dataset and trying to implement personalized plans. It has been used in healthcare to group at-risk patients and recommend diet or exercise plans for the clusters or groups (Babcock, 2016; *Predictive Analytics Models and Algorithms*, 2022).

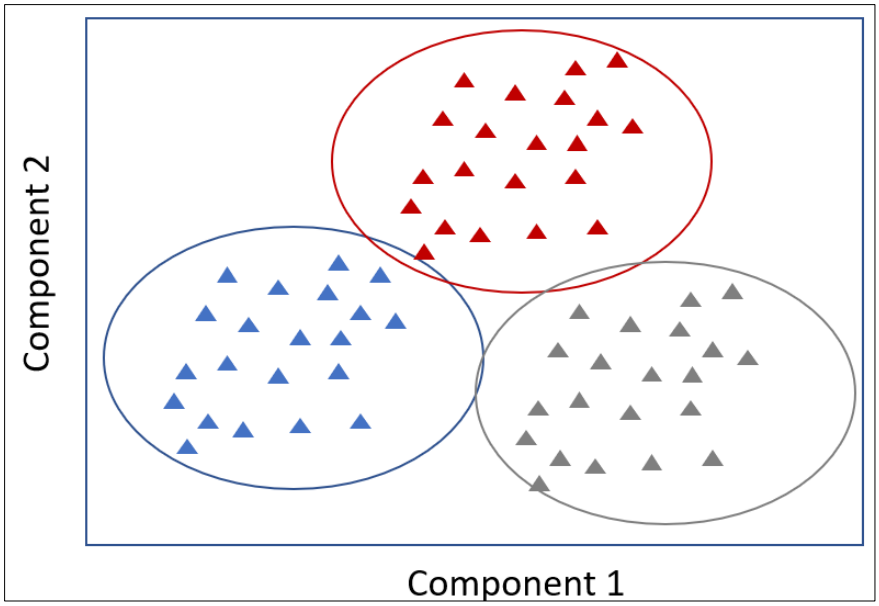


Figure 2.4: Illustration of K Means Algorithm

Note: From Predictive Analytics Models and Algorithms

2.2.2.3 Prophet

The prophet algorithm is an open-source algorithm developed by Facebook, it is used in time series and forecast models. It is of great use when it comes to capacity planning. Forecasting models tend to be inconsistent and inflexible, manual forecasting is not exactly an option as it is labour intensive both in terms of skill and manhours. Prophet is a successful automation of forecasting that can work with useful assumptions. It has gained popularity due to its speed, reliability and robustness while dealing with "messy" data (*Predictive Analytics Models and Algorithms*, 2022).

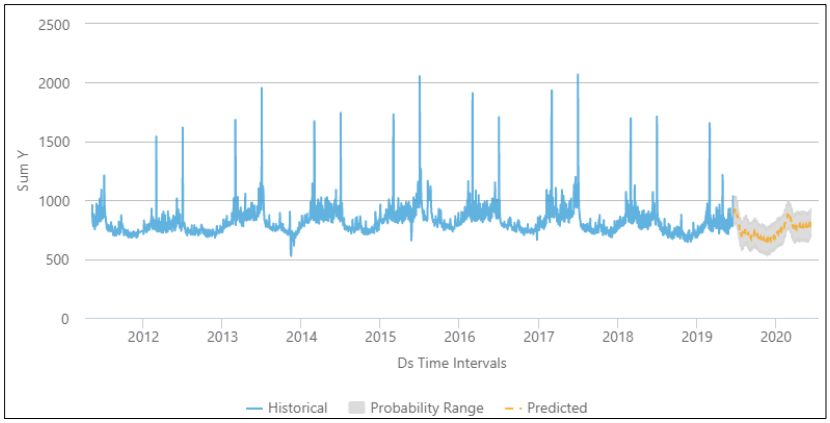


Figure 2.5: Illustration of Prophet Algorithm Output

Note: From Predictive Analytics Models and Algorithms

2.3 Diabetes Management

2.3.1 Diabetes Overview

Diabetes is a chronic disease caused by a lack of insulin production by the pancreas or by poor utilization of the insulin that is produced. (IDF, 2020). Insulin is a hormone that helps glucose get to blood cells and produce energy.

There are 3 types of diabetes:

- Type I Diabetes

  Mostly occurs in children and adolescents and is caused by the pancreas producing little or no insulin. It has no cure which means daily insulin injections are needed to manage it.

- Type II Diabetes

  It accounts for 90% of diabetes cases and mostly occurs in adults. With type II diabetes the body does not utilize the produced insulin efficiently. The treatment for type II diabetes is a healthy diet and increased physical activity along with medical management.

- Gestational Diabetes

  This is characterized by elevated glucose levels during pregnancy, it is often temporary and reverses after childbirth. However, in most cases, the risk of getting type II diabetes increases after gestational diabetes.

According to IDF, 537 million adults are living with diabetes globally. The report projects that these numbers will rise to 634 million by 2030 and 783 by 2045 for patients aged 20 to 79 years. It also places the number of people living with undiagnosed diabetes at 240 million. The IDF diabetes atlas places 75% of people with diabetes in middle- and low-income countries (IDF, 2021).

In Africa, the current regional prevalence of diabetes stands at 4.5% and is expected to rise to 5.2% in 2045, translating to 1.05 billion with diabetes. Kenya is at par with the rest of Africa, with higher rates found in urban areas, with a marked increase in obesity, poor dietary habits, excessive alcohol consumption and inactivity (Jones, 2013).

The level of knowledge about diabetes and its management via nutrition is low, at about 30%. Trials across different regions have indicated that a healthy diet can delay or altogether prevent the onset of type II diabetes (Frost et al., 2003).

2.3.2 Nutritional Management of Diabetes

People are often not aware of the nutritional deficiencies in their bodies, especially in this era of processed foods, which increases the cases of obesity and related diseases such as type II diabetes. In the event that one undergoes nutritional testing, deciphering the results can prove challenging without access to a nutritionist or dietician. In Kenya, the service ratio is 1:200000. (Kenya National Bureau of Statistics, 2019b), which gets even lower away from urban centres. Medical Nutritional Therapy (MNT) is recommended in helping to combat type II diabetes and reduce reliance on medication thereby reducing the cost of diabetes care. Macronutrients provide energy for the human body while micronutrients are the organic compounds, vitamins and minerals, needed in small amounts for normal processes of the body (Shafer, n.d.). Both can be obtained through a varied and balanced diet.

Nutritional testing helps identify which nutrients are deficient allowing patients to tailor their diets for a healthier lifestyle. The insights from nutritional testing aid in functional medicine such as blood sugar regulation, weight management, and organ function just to name a few. Since nutrients play a key role in most bodily processes, imbalance over time can have symptoms ranging from weight gain to life-threatening diseases. Diet modification is an easy way to eliminate deficiencies and achieve optimal health (*Benefits of Blood Testing for Nutritional Deficiencies*, n.d.).

2.4 Existing Predictive Tools

Predictive tools are used in an array of fields such as:

- Finance

Using data from the industry and previous financial statements, it is possible to predict future earnings, sales and expenses that can be used in decision making. Tools in this space included Microsoft Azure's Power BI which also offers visualization of trends. (Aspin, 2020; Eckerson, 2007).

Commented [M6]: Is this tool documented in 2007 still used?

Commented [AK7R6]: Added further referencing

- Marketing

In retail, there is an abundance of consumer data which can be leveraged for targeted marketing strategies to better reach potential and return customers (Eckerson, 2007).

- Entertainment & Hospitality

Predictive analytics can be employed to mitigate staffing challenges in hospitality. In entertainment, the tools used are recommender systems which are a subset of predictive analytic tools which provide a set of recommended outputs that optimize a beneficial outcome (Falk, 2019).

- Healthcare

Similar to entertainment, recommender systems are widely used to offer dietary and lifestyle recommendations in healthcare. They are also used in other scenarios such as allergy detection with the AbbieSense sensor or sepsis detection (Nithya & Ilango, 2017; Teng & Wilcox, 2020).

- Manufacturing

In manufacturing, the tools can be used to predict when a malfunction is likely to occur, for logistical planning as well as supply chain and inventory management which is what SAS Predictive Analytics offers (Eckerson, 2007; Shmueli & Koppius, 2011).

- Welfare

In child welfare cases, predictive analytics tools can be used to determine the level of endangerment or to recommend what steps can be taken to ensure a child's overall well-being such as parental support. This allows the agencies in question to better prioritise cases while the recommender systems can offer a case-by-case to-do list of training and services that can be offered to affected families.

Below are some existing predictive tools in healthcare and nutrition

2.4.1 Predictive tools in healthcare

Nithya & Ilango, 2017 studied the applications of predictive analytics across different areas of healthcare. The use of predictive analytics in electronic record management, disease prediction and computer-aided diagnosis were found to offer risk management tools leading to improved patient safety and healthcare quality while reducing healthcare costs and providing personalized care (Boukenze et al., 2016; Nithya & Ilango, 2017).

2.4.2 A Novel Software to Improve Healthcare Based on Predictive Analytics and Mobile Services for Cloud Data Centers

This software aimed to identify patients who were likely to be admitted or readmitted to the hospital by using historical claims data including patient data. This would enable hospitals and insurance companies to better plan for the coming business year. It worked by using predictors based on time, accuracy and the cost of the hospital stay given certain patient parameters (Shmueli & Koppius, 2011).

2.4.3 Predictive Analytics Solutions for Sepsis Detection

This study by Teng & Wilcox reviewed several existing techniques for sepsis management with various techniques ranging from linear regression, logistic regression, support vector machine (SVM) and the Markov model. The reviewed tools required vital signs and pathological tests to help them detect sepsis and in some cases death by sepsis. While some of the techniques reviewed were theoretical they showed promise for real-world application which could be beneficial and even life-saving to patients if sepsis were detected early enough (Teng & Wilcox, 2020).

2.4.4 Market2Dish: Health-aware Food Recommendation

Wang et al., (2021) proposed a health-aware recommendation system that maps ingredients in the market to possible healthy homemade recipes. The system had three components: recipe retrieval, health profiling and finally the food recommendation module.

It was implemented by using a novel category aware hierarchical memory-based network, the network allowed it to learn from the user-recipe interactions to improve the recommendations offered.

2.4.5 DIETOS: A novel food recommender system

DIETOS is a web-based recommender system by Agapito et al., (2016) for the recommendation of nutrition content to improve the quality of life for people living with diabetes and even those without.

The system constructs a health profile and provides nutritional recommendations as per that profile. The system has several independent modules that interact via communication interfaces.

2.4.6 Diet-right: A smart food recommendation system

Diet-right is a smart food recommendation system that recommends foods and nutritional changes based on the user's pathological test reports. From the report data, the system constructs a user profile, and any abnormality levels are recorded.

Foods are recommended based on the recorded abnormalities. It uses the Ant Colony Optimization technique which is a population-based approach to determine which food items to recommend (Rehman et al., 2017).

2.5 Existing Avenues to Access Diabetes Nutritional Information

An analysis of nutrition dietary patterns led by Serge Hercberg observes that to combat most NCDs and achieve a sustainable public healthcare system, there must an understanding of the relationship between nutrition and health. While in some regions healthcare systems are sufficiently serving the populace now, they will eventually feel the strain of increased demand with limited resources. Ideal healthcare systems should be sustainable, economical, and reliable (Faezipour & Ferreira, 2011). Currently, there is a minimal understanding of the different nutrition sources in the country leading to over-reliance on some resources to produce specific products while ignoring other possible options.

This results in over-exploitation of the environment and nutritional deficiencies across the food supply chain. If the populace had access to a reliable body of information regarding best nutritional practices and the relationship between nutrition and health, some of the challenges with diabetes management in Kenya would be addressed. Access to this information would help with better health choices (Hercberg et al., 2010).

2.6 Summary of literature review

Table 2.1: Summary of Existing Predictive Tools

System	Description	Techniques Used	Gaps
Market2Dish	Health-aware Food Recommendation	Category aware hierarchical memory-based network	Not nutrition aware, recipe based.
DIETOS	A novel food recommender system	Independent PHP modules with communication interfaces	Disease aware, not nutrition aware
Diet-right	A smart food recommendation system	Ant Colony Optimization	Nutrition aware, not tailored to Kenya and not location aware
Sepsis Detection	Sepsis detection tools	SVM, Linear Regression, Logistical Regression	Used pathological tests to identify sepsis levels, not nutrition-based

2.6 Identified Gaps

Most of the predictive tools and systems in the related works focused on recipe retrieval, user preference-based food recommendation, cooking assistance, or the nutrition and calorie estimation of dishes. Not many of the systems focused on user nutritional markers and none offered location aware predictions.

This is a gap that can be used to improve the diet and overall health of diabetic patients. The prototype designed for this study predicts foods not only based on nutritional deficiencies but also on the provided location. The model offers predictions of foods with high quantities of the deficient micro or macronutrients as per the FAO food classification tables (FAO/GOK, 2018).

2.7 Conceptual Framework

Commented [AK8]: Updated illustration to include the user (external feedback)

Based on the literature review and the gaps that were subsequently identified, this study proposed a conceptual model with a two-pronged approach. Using the patient's nutritional markers to determine any deficiencies in macronutrients and micronutrients according to the Laboratory Test Reference Ranges (ABIM, 2022). For the identified deficiencies, the tool returns predictions of food items from the food composition tables for Kenya and then further cross-references those with location-based food items. The nutritional needs prediction tool will have a feedback module built in to enable a nutritionist to review past predictions and eliminate any invalid predictions.

Commented [VO9]: typos

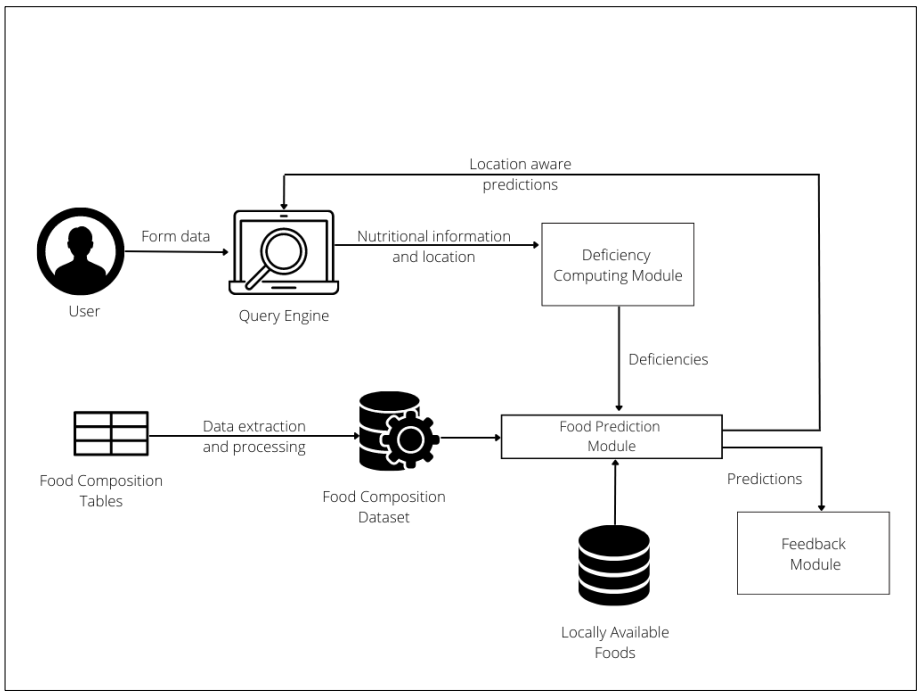


Figure 2.6: Nutritional Needs Prediction Tool Conceptual Framework

Note: Illustration of how the predictive tool components integrate.

Chapter 3: Research Design and Methodology

3.1 Introduction

This chapter outlines the research design, data collection, development of a location-aware predictive tool interfaced by a web-based application that is accessible across multiple devices. It includes the steps and procedures that were taken to develop the solution. The methodology used is structured system design methodology which breaks the system down into functional modules, with each module having inputs, processing, and outputs. Structured system design is a top-down decomposition of the system's functionality (Guthrie, 2003). This study determined the relationship between two sets of variables, the dependent variables which are the patient's nutritional profile and location, and the independent variable which are the food items suggested.

Commented [VO10]: Poor sentence construction

3.2 Research Design

Research design is defined as a plan for a study, providing the overall framework for collecting data (Leedy et al., 1997). Blanche et al. define research design as a strategic framework for action that serves as a bridge between research questions and the execution, or implementation of the research strategy (Blanche et al., 2006).

Research design allows for the researcher to focus on the most suitable research methods. The problem is what defines the design and not vice versa. This phase also helps identify the tools needed.

The study used quasi-experimental research design which aims to establish a cause-and-effect relationship between an independent and dependent variable without random assignment. In this case, the dependent variables are the patient's nutritional profile and location, and the independent variables which are the predictions of food items.

3.3 Software Development Methodology

Agile software development methodology employs continuous planning, learning, and improvement and encourages flexible responses to change. It promotes continuous iteration of development and testing throughout the software development lifecycle of the project with development and testing occurring concurrently to allow for the constant application of fixes as issues are identified (Palat & Hastie, 2021). This flexibility and the iterations in the agile SDLC were ideal for this study as they helped ensure the accuracy of the predictive tool given

the uniqueness of the input data i.e., the patient's deficiency needs and location. Figure 3.1 below illustrates the phases of agile SDLC:

i. Planning

   This is the initial phases where the objectives of the study were reviewed to determine the technical feasibility. It is also when the resources required to build the system were identified.

ii. Analysis

   In the analysis phases we defined the system requirements, both functional and non-functional. This also involved defining more detailed elements of the system such as inputs, processes and outputs which were captured in a system architecture diagram.

iii. Design

   This phase involved creating mockups of the application that the patients and nutritionists would interact with. Using various UML diagrams the features of the system were also captured to ensure the requirements were all met.

iv. Implementation

   In this phase, the design documentation and diagrams were converted into code to result in a working product. It involved iteratively developing and testing the code to ensure the outputs were as defined in the system requirements.

v. Testing and Integration

   The testing of the system included both unit testing for the code and user testing to ensure the objectives of the system were met.

vi. Maintenance

   The maintenance stage was not yet needed for this study but would involve ensuring the packages used were up to date and any changes required to the system features are made.

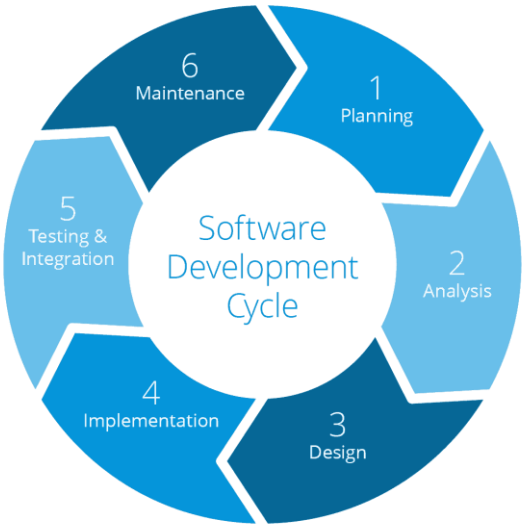


Figure 3.1: Phases of Agile Software Development Lifecycle

Note: From *Agile Methodology*, n.d (https://www.javatpoint.com/software-engineering-agile-model)

3.3.1 System Analysis

Commented [VO11]: Rearrange to have Research Design as section 3.2
And Software Development Methodology as section 3.3.

The current section 3.3.1 and 3.3.2 fit under Software Development Methodology and not Research Design

The agile system analysis process consists of 5 steps which were followed in this study

i. Identify System Users

   This is an important step since if the users are wrongly identified, then the solution built will be wrong. The users for the predictive tool were identified as the patient and the nutritionist.

ii. Define Main User Goals

   The users in the system have specific goals. For the patient, they will use the system to get suggestions on food items that can address deficiency needs having provided their location and nutritional profile. The nutritionist can use the system in a similar manner as the patient and they can also use the system to provide feedback on previously suggested food items for specific nutritional deficiencies.

iii. Define System Usage Patterns and Functional Solution

   Each user of the system has behavioural patterns, such as login in to provide feedback as a nutritionist. This helped get started with identifying functional solutions by clearly

outlining the problem and working towards a simple, effective, and elegant solution by identifying the best way to satisfy the usage patterns.

iv. Define Main Navigation Paths

Navigation paths concentrate on the information needed by users on each step, the navigation paths identified for this system were:

- Patient
  - Add nutritional results
  - Add biomarkers
  - View BMI and status
  - View deficiencies
  - View general food predictions
  - View location-aware food predictions
- Nutritionist
  - Log in to the system
  - View past predictions
  - Validate past predictions

v. Create UI Mockups

This was the final step of system analysis, the creation of UI mockups which illustrated the defined navigation paths.

This process provided a blueprint for the system implementation and design and helped with the determination of functional and non-functional requirements.

3.3.2 System Design

The system design was according to the output of the planning and analysis phases. It was designed to meet the resultant system requirements.

Software modelling diagrams were used to show the structure and logical interactions of the various component of the predictive tool which is accessed via a web-based application.

i. Context Diagram: To define the boundary between the system, its environment and the entities that interact with it.

ii. Data Flow Diagrams: To provide a simplified representation of the flow of data in the system, including inputs, outputs, data stores and processes that manipulate/consume the data.

iii. Entity Relationship Diagrams: To provide a graphical representation of the actors, objects, concepts, or events with the system.

iv. Database Schema: To represent the storage of data in our dataset and application database, covering the data types, the organization of the data and relationships between tables in the data store.

3.4. System Implementation

Commented [VO12]: Numbering of sub- section under section 3.4 is haphazard

The design was converted into a prototype by coding it. Agile methodology allows for iterative development which fine-tunes both the predictive tool and the application.

3.4.1 Predictive Tool Implementation

These are the steps that were taken to populate and build the predictive tool:

i. Data extraction and collation

ii. Identifying a measure of success of the predictive tool

iii. Develop a location-aware nutritional needs predictive tool

iv. Validate the predictive tool

3.4.1.1 Predictive Tool Development

There are many libraries for model development and data analysis. Pandas is one such library, a fast, powerful, flexible and easy to use open-source data analysis and manipulation tool, built on top of the Python programming language.

3.4.1.2 Predictive Tool Validation

The predictive tool was evaluated for accuracy once completed. The formula used for computing accuracy was

$$percentage\ accuracy = \frac{correct\_predictions}{total\_predictions} \times 100$$

3.4.2 Implementation environment

Due to the costly nature of GPUs, the logical approach was to work with GPUs on cloud. The cloud environment provides:

i. High performance with the latest technology.

ii. Cost-effective, one only pays for the resources that are used as is the norm with PaaS. In this case, Google's Colab was used as it offers a free trial quota.

3.4.3 Programming tools

For the model and application construction, Python was used as it has the most AI/ML libraries including those needed for data analytics. The library Pandas was used for its implementation of matrix factorization. For the web-based application, Python for the backend with the Flask framework and HTML/CSS for the web application.

3.5 System Testing

Reliability and validity are often the concepts used to determine research quality. Reliability is the consistency of data and validity is the accuracy of data. The reliability of the tool was checked, i.e., if given the same or similar data the output should be consistent. Similarly, validity was checked by focusing on the percentage of successful approximations, meaning for a given user query, was the resultant information relevant? In addition, there were unit and functionality tests for the platforms that were integrated with the prediction tool.

3.6 Target Population and Sampling

A target population generally refers to a large collection of individuals or objects that is the main focus of a scientific query or research (*Research Population*, 2009). For the purposes of this study, the target population is the population of Kenya. The sampling will be clustered by region, i.e., rural vs urban. It will be a two-stage cluster sample, from the regional clusters the study will mainly focus on individuals above the age of 15.

3.7 Data Collection

The data for this study was retrieved from two main sources:

- FAO in association with the government of Kenya for the Food Composition tables.
- The American Board of Internal Medicine for Test Reference Ranges

3.8 Data Analysis

3.8.1 Data Cleaning

The data collated from the FCT had attributes that were not required for the system, these were removed as part of the cleaning process. The data was also checked for correctness, ensuring that the compositional values were all numerical (float or int). In the case of missing attributes, zero imputation was applied. The handling of missing data ensured that the dataset is accurate and contains valid information.

3.8.2 Data Grouping

The data was grouped by food items and by location. For each location (county) the available food codes were mapped to allow for cross-referencing of predictions thereby returning location-aware predictions.

The feedback module queued the food item predictions and related nutrients by eliminating any invalid or impractical predictions.

3.9 Research Quality

This study aimed to build on top of existing predictive analytics tools, algorithms and machine learning techniques. This research did not re-invent the wheel on artificial intelligence and machine learning. The implementation is well documented to ensure that the study can be adopted and replicated by others in similar domain structured data sets. A small test data sample was used to determine the accuracy of the model in retrieving the relevant information.

3.10 Ethical Approval

The necessary approval to conduct this research was acquired from the Ethical Committee at Strathmore University. This was to ensure that the data used in the study was acquired legally and as per the ethical guidelines. Additionally, this study used publicly availed data on nutritional management and the prevention of diabetes. The researcher will disclose to the owners the intent to use the data. Any previous research referenced during this study has been duly acknowledged. This includes citation and referencing of the source of the content, tables, and images.

Chapter 4: System Analysis, Designs, and Architecture

4.1 Introduction

This study sought to develop a prediction tool based on location and nutritional markers for diabetic patients in Kenya. The objectives of the study were to fill the gap in public awareness when it comes to nutritional management and prevention of type II diabetes by making the prediction too accessible to the general public.

This section outlines the system design and architecture of the implemented solution. It includes the various Unified Modelling Language (UML) diagrams that guided the design of the system; use case, sequence, data flow, context, and entity-relationship diagrams to visually present the solution. The wireframes used for the web application search engine are also captured in this chapter.

4.2 Requirements Gathering and Analysis

The general requirements gathered for this study were obtained from two sources, the Kenya National Bureau of Statistics (KNBS) and the Global Nutrition Report. The fourth volume of the 2019 national census report by KNBS and the country profile from the 2021 Global Nutrition Report were analysed.

4.2.1 Analysis

The census report contained tabular data showing the distribution of the population by socio-economic characteristics. For the purposes of this study, we focused on summary findings around information and communication technology (ICT).

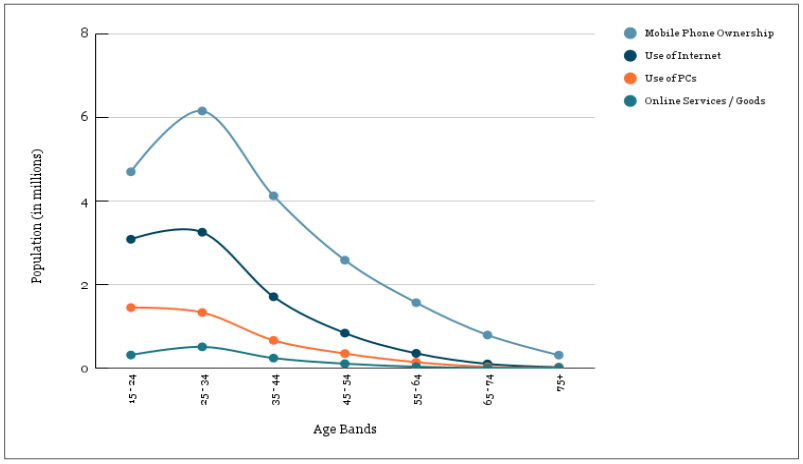


Figure 4.1: Distribution of Population Age 15 years and above who owned and used Selected ICT Equipment and Services by age

Note: Derived from Appendix C: Table C.1

Please see Appendix C: Table C.1 which was analyzed resulting in Figure 4.1 shows the use and ownership of certain ICT equipment and devices according to age. The data analyzed is limited to those aged 15years and above. Table 4.2 below was derived from the contents of Appendix C: Table C.1; the values were obtained using the formula below:

$$Percentage\ Use = (Subset\ Use / Total\ Population) * 100$$

Table 4.1: ICT penetration in Kenya based on ownership and use of selected ICT equipment and services

Mobile Phone Ownership	Use of Internet	Use of PCs	Use of Online Services
70.48%	32.71%	13.97%	4.35%

Note: Derived from Appendix C: Table C.1

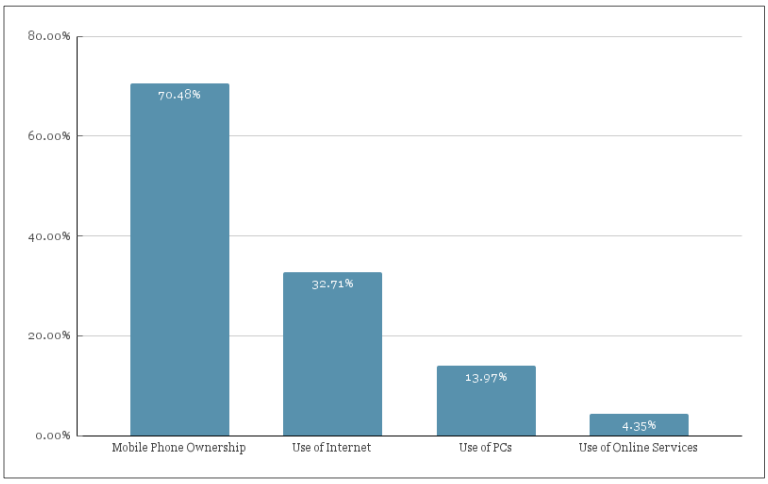


Figure 4.2: ICT penetration in Kenya based on ownership and use of selected ICT equipment and services

Note: Derived from Appendix C: Table C.1

The data was also analyzed based on region i.e., Rural vs Urban. The results were split into two broad categories:

- Population aged 3 years and above who owned mobile phones, used the internet, or used personal computers (laptops, tablets, and desktop computers)

- Population aged 15 years and above who searched for and bought goods and services online

Appendix C: Tables C.2 and C.3 contain the data as extracted from the census report, while table 4.4, figure 4.3 and figure 4.4 were derived from that data to better illustrate the availability of these resources nationwide.

Table 4.2: ICT penetration nationally based on ownership and use of selected ICT equipment and services

Mobile Phone Ownership	Use of Internet	Use of PCs
47.31%	22.57%	10.35%

Note: Derived from Appendix C: Table C.3

Figure 4.3 below shows a further breakdown of the use of ICT equipment and services. For instance, 59.19% of the 47.31% of the population who own a mobile phone are from rural areas, and the remaining 40.81% are from urban areas.

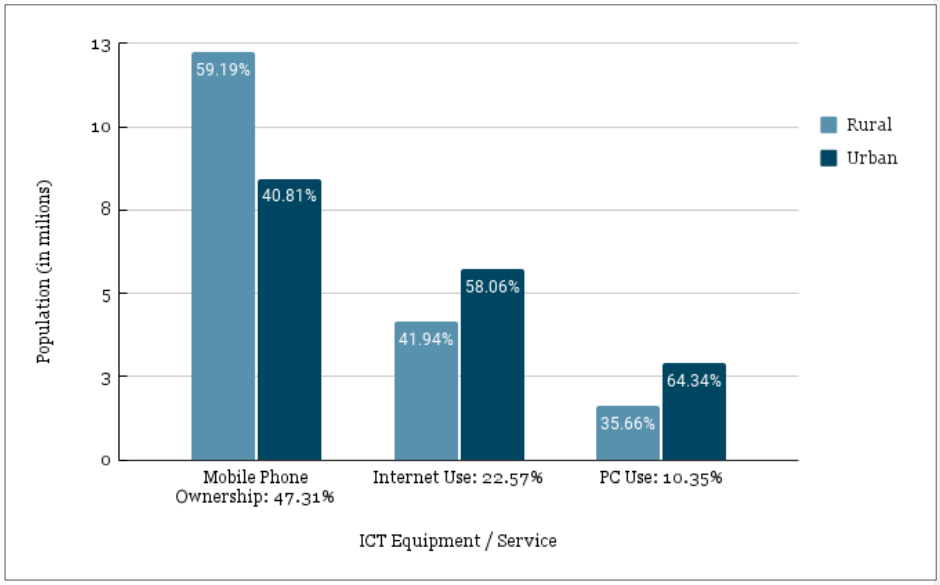


Figure 4.3: Distribution of Population Age 3 years and above who owned and used Selected ICT Equipment and Services by region

Note: Derived from Appendix C: Table C.3

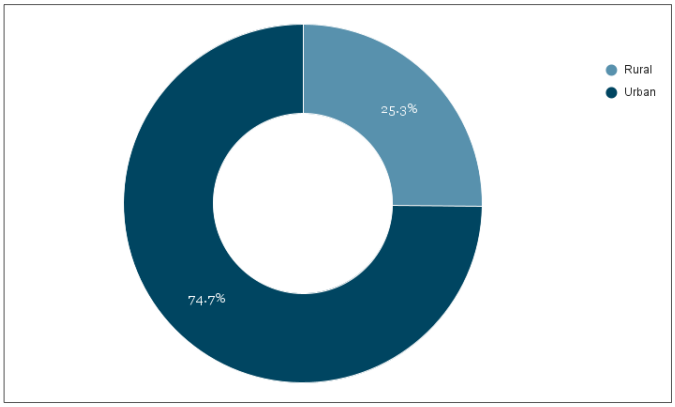


Figure 4.4: Distribution of Population age 15 years and above who Searched and Bought Goods and Services Online by region

Note: Derived from Appendix C: Table C.3

Kenya's country profile analyses the country's progress towards the global nutrition goal on diet-related non-communicable diseases (NCD) including diabetes. The analysis report on diabetes is based on age-standardised modelled estimates for adults aged 18 years and older, using the WHO standard population and they are reported by sex. The report was available on the website, along with the raw numerical data and the methodology used to analyse it.

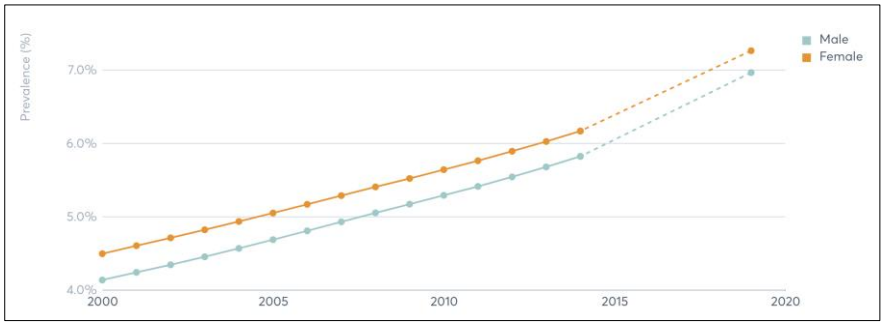


Figure 4.5: Prevalence of diabetes in Kenya based on gender

Note: Extracted from the 2021 Global Nutrition Report, Country Profile – Kenya

The data indicating mortality attributable to dietary composition and weight was extracted from the raw data and analyzed in two broad categories.

- Mortality due to weight-related issues, i.e., being obese or overweight

- Mortality due to diet only, i.e., of acceptable weight but still attributed to dietary composition

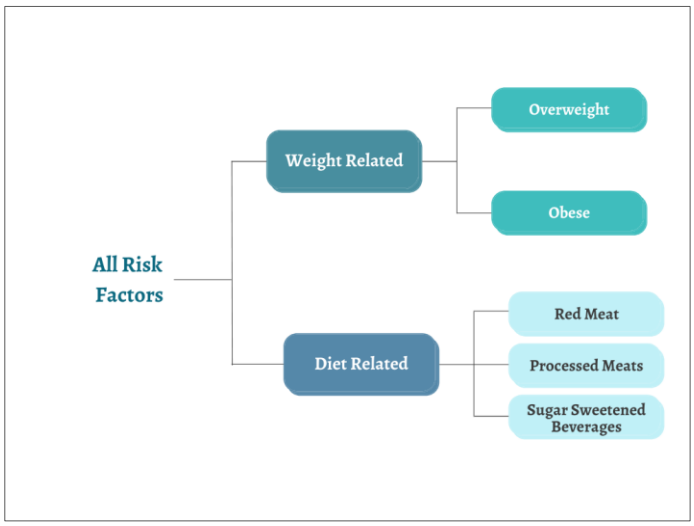


Figure 4.6: Illustration of categories of dietary and weight-related risk factors for diabetes

Note: Extracted from the 2021 Global Nutrition Report, Country Profile – Kenya

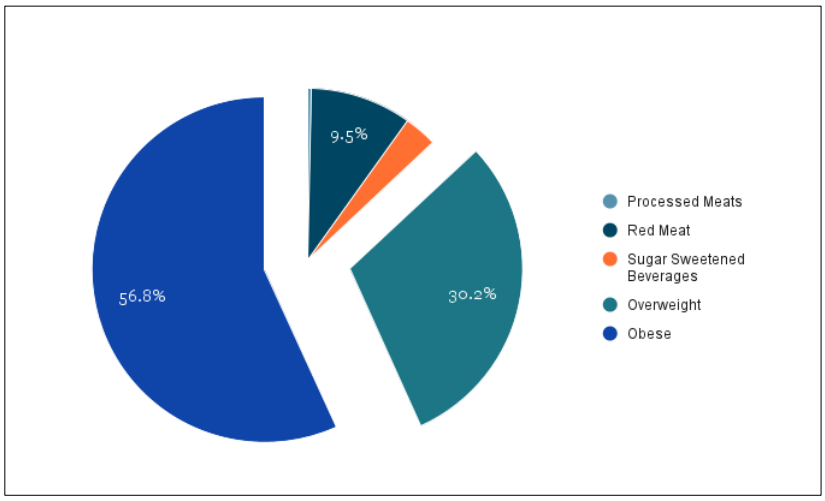


Figure 4.7: Mortality attributable to dietary composition and weight

Note: Derived from the 2021 Global Nutrition Report, Country Profile - Kenya

4.2.2 Requirements

The features and constraints of the location-aware nutritional needs prediction tool can be split into functional and non-functional requirements:

4.2.2.1 Functional Requirements

For the system to achieve the objectives of this study, it needed to have some capabilities and perform certain functions and processes:

- The user should be able to query the system for nutritional predictions based on context. The context will include both primary context (location) which is required and secondary context (biological markers) which will be optional.
- The validator (nutritionist) should be able to rate the accuracy of results previously returned by the system to user queries on nutritional information.
- The system should process the nutritionist's feedback and use it to improve the accuracy of the information it responds with to user queries.

4.2.2.2 Non-Functional Requirements

Additionally, the system met these requirements. While the system could have functioned without meeting them, they make the system interactive and user friendly:

- Security: The administrator and validator should be authenticated (username and password) and authorized to view system data.
- Storage: The system should be capable of storing the various datasets.
- Performance: The response time to the user queries should be acceptable and the system should be able to handle multiple concurrent users.
- Scalability: The system should allow for functionality changes as requested by the user over time.
- Usability: The system's users should be able to query the system and review the feedback provided.

4.3 System Design

System design is the process of defining the components, modules, interfaces, and data for a system to satisfy specified requirements (Blanchard & Fabrycky, 2010). This used the CPMAI methodology which allows for iterative and low-risk development and fine-tuning of the system. The system is a location-aware nutritional needs predictive tool with a web page to

facilitate querying, making it well suited for agile methodology which is iterative allowing for constant fine-tuning and improvement.

4.3.1 System Architecture

The food composition tables were gotten from the FAO website and the relevant tables were extracted and cleaned to result in a functional data set. The web application provides an interface for the querying of the data set as well as validation of previous results from the system. Figure 4.8 shows an overview of the system.

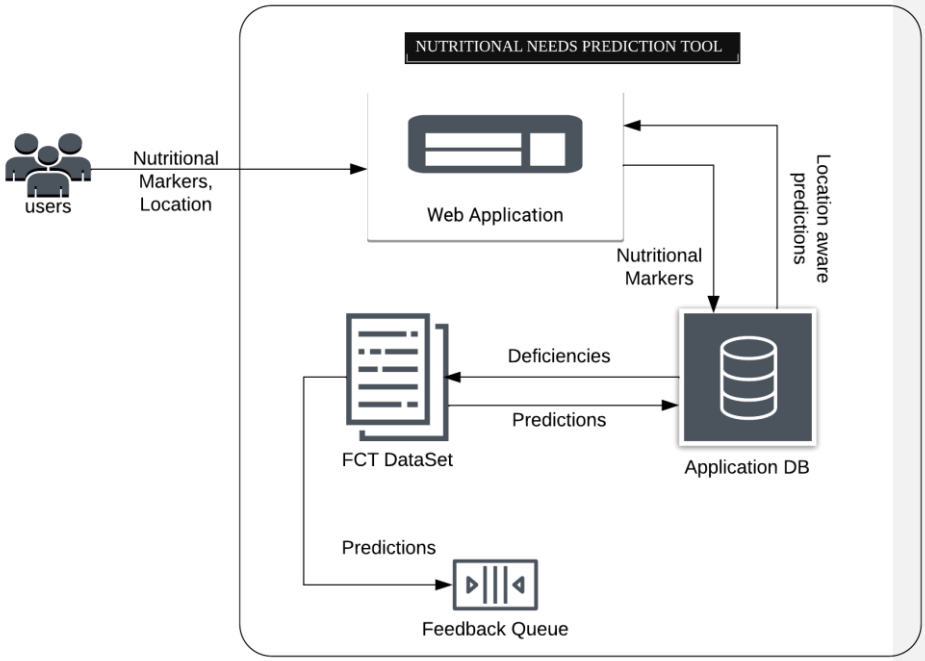


Figure 4.8: System Architecture

4.3.2 Context Diagram

The context diagram below, figure 4.9, shows the relationship between the system and external entities. These entities include the user, system administrator, and approved nutritionists. The users query the system for nutritional information. The nutritionists validate the responses

previously returned by the system. The system administrator ensures the availability and functionality of the system.

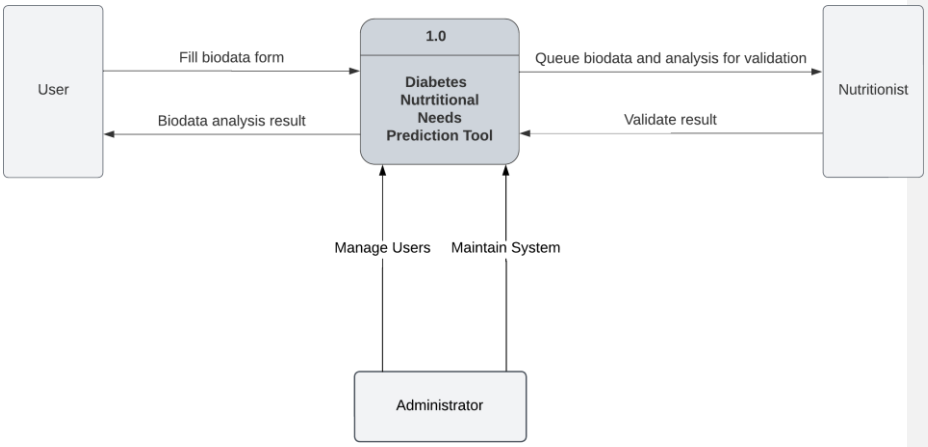


Figure 4.9: Context Diagram

4.3.3 Data Flow Diagram

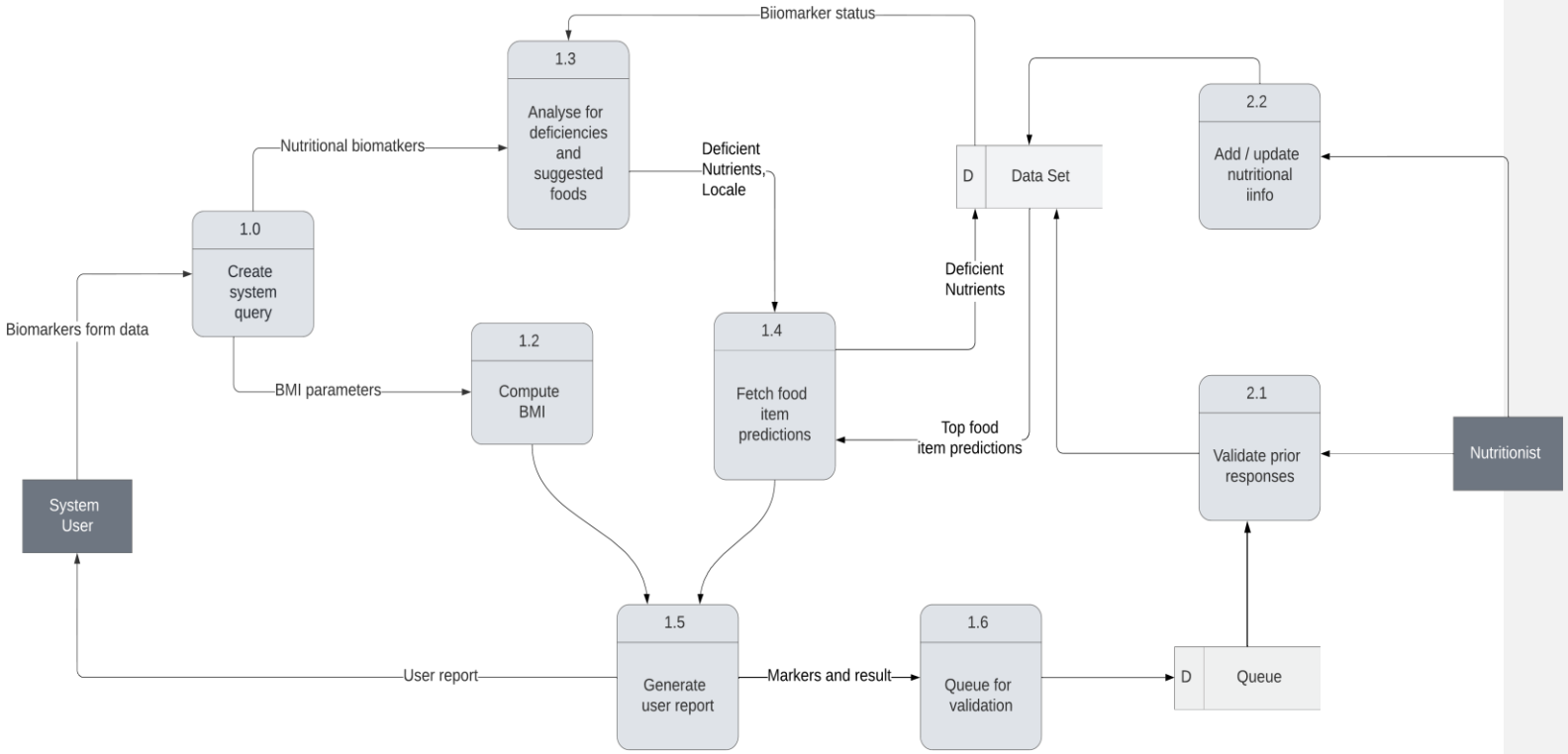


Figure 4.10: Data Flow Diagram

4.3.4 Database Schema

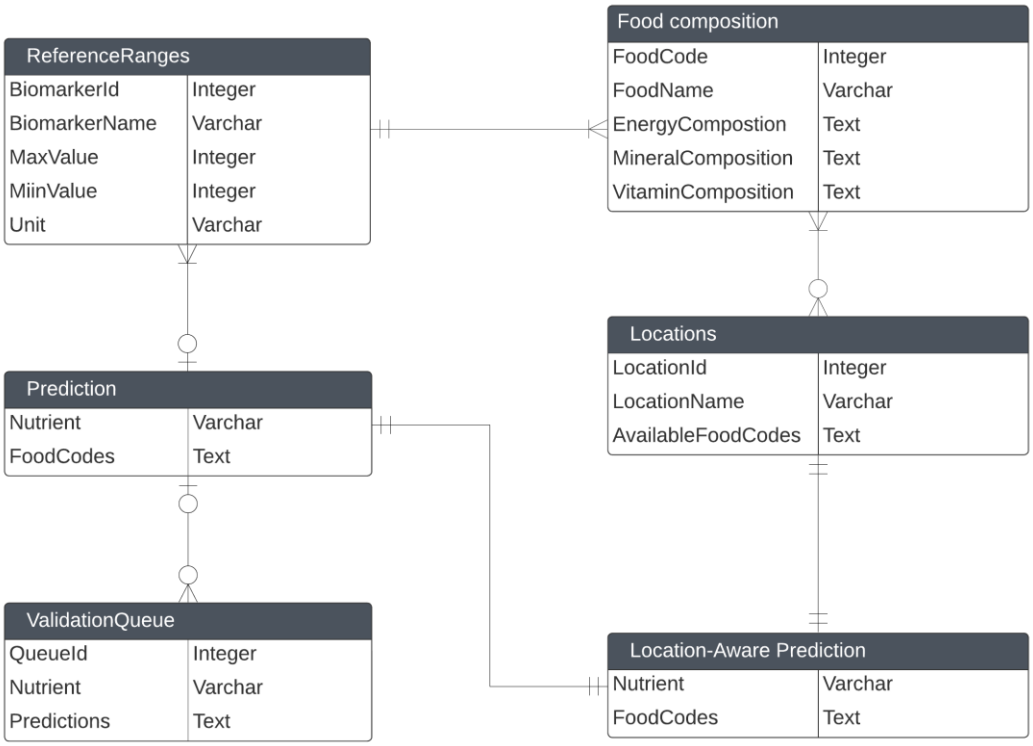


Figure 4.11: Database Schema

4.3.5 Entity Relationship Diagram

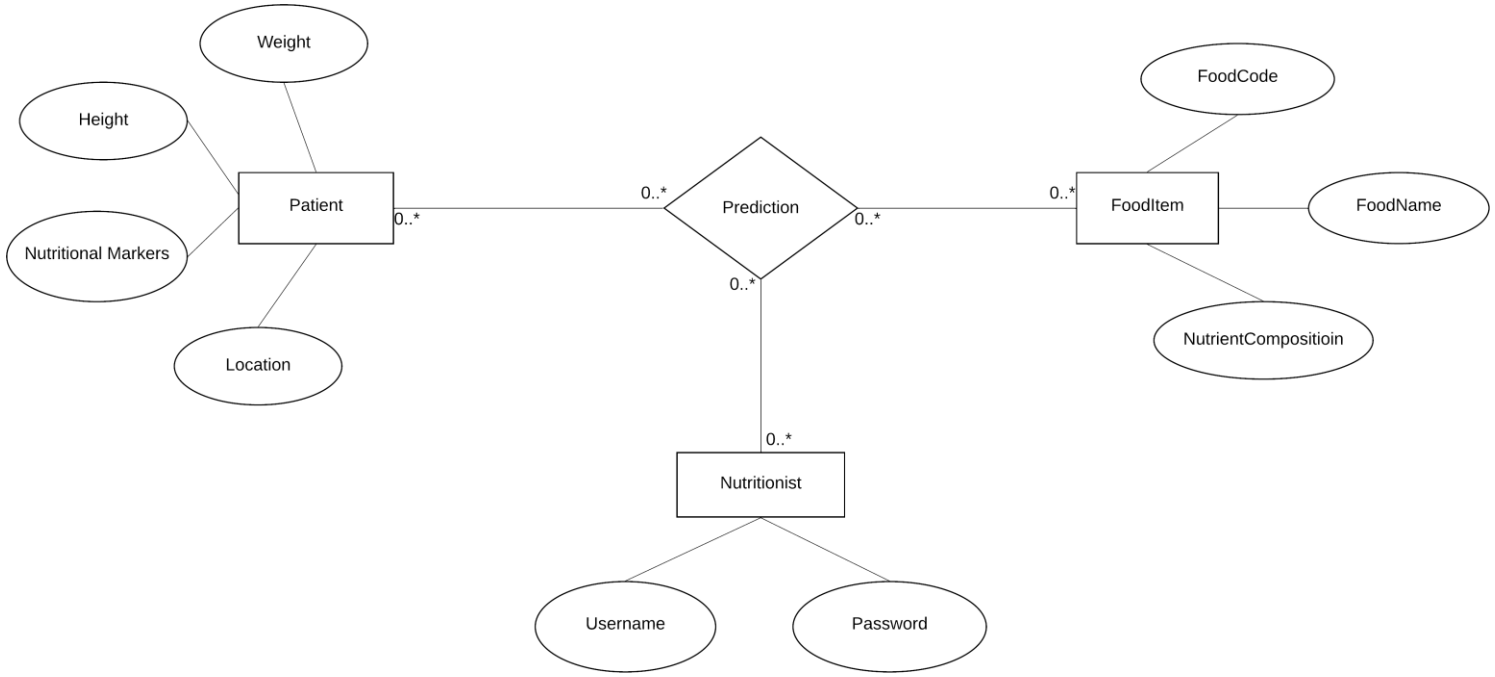


Figure 4.12: Entity Relationship Diagram

4.4 Web Application Wireframes

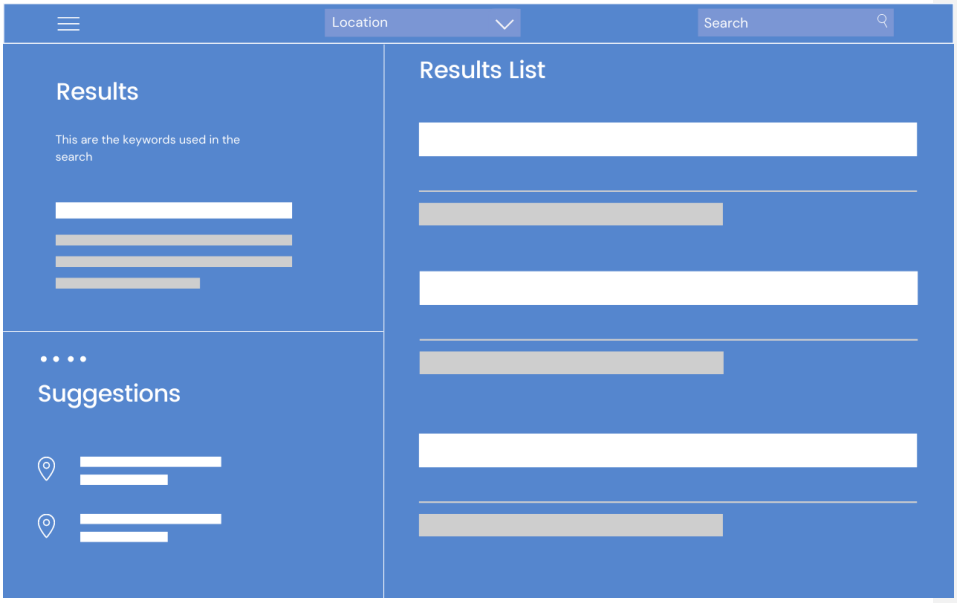


Figure 4.13: Search Page

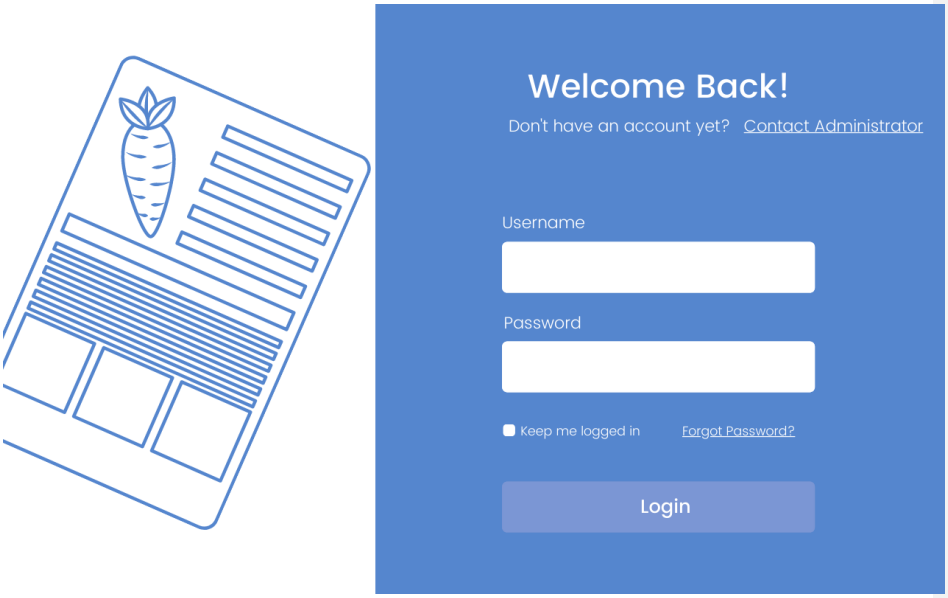


Figure 4.14: Login Page (Validators Only)

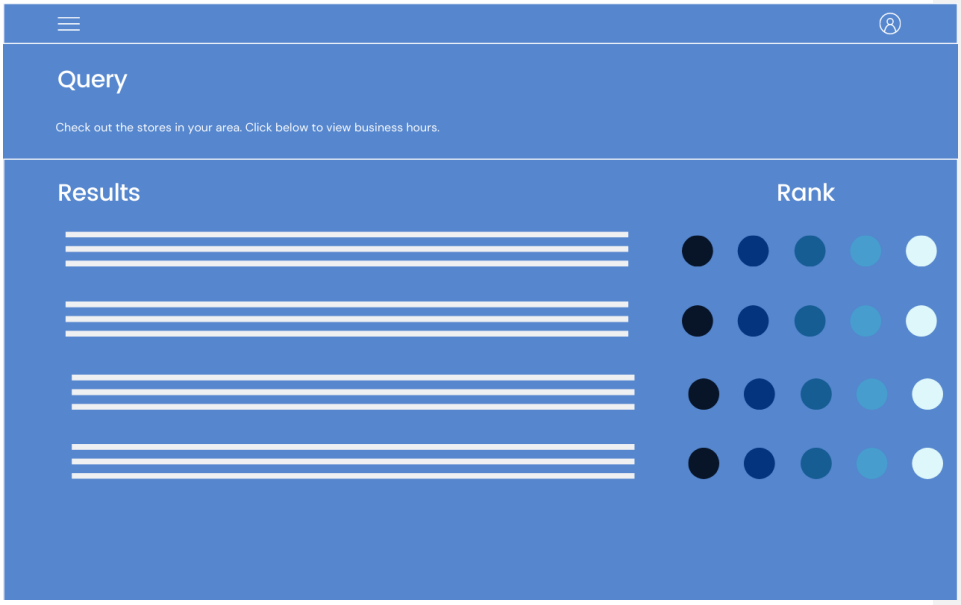


Figure 4.15: Validation Page (Validators Only)

Chapter 5: System Implementation and Testing

5.1 Introduction

The section focuses on the development of the prototype of the location-aware nutrition needs prediction tool. It covers the building, testing, and validation of the prediction tool. The implementation phase involved the development of the various modules of the resultant application. The testing and validation phases focused on functional and usability testing to determine whether the system met its objectives.

5.2. System Implementation

Agile methodology was used in this study to allow parallel development of the data analysis module and the actual web application. The iterative capabilities allowed modular modifications to be made to fine-tune the system to ensure the study's objectives were met. The prediction tool is part of a responsive web application that was developed using python, specifically the Flask framework. The templates rendered for user interaction were developed using HTML and CSS for styling. The database used was PostgreSQL to store, food name mapping and food availability by counties in Kenya.

The datasets containing the food composition tables were analysed using Pandas, a python data analysis and manipulation tool. For the predictive functionality, Panda's matrix factorization implementation was utilized. Matrix factorization is a mathematical model that splits an entity into smaller entities, through an ordered array of numbers. In this case the entity being split is the food composition table for Kenya, and it is split by nutrients into smaller entities which are then ordered by decreasing quantity for each food item.

There was also a feedback module built into the prediction tool that leveraged CloudAMQP which is a SaaS offering of cloud hosted clusters of RabbitMQ, an open-source message broker. To use the application, the patient requires the results from nutritional testing. This is fed into the web application along with certain biomarkers and a county location selection. The application computes deficiencies using the laboratory reference ranges. For the computed deficiencies, it uses the prediction tool to predict nutrient heavy food items and further cross-references them with the food items available in the location selected.

The predictions from the food composition dataset are queued to the CloudAMQP instance. When a nutritionist logs in to the system, the predictions are retrieved from the queue in batches and the nutritionist eliminates the invalid predictions to ensure they are not predicted for that particular nutrient in the future.

5.2.1 Hardware requirements

The hardware system requirements for developing the web application were:

i. A laptop with Windows 10 OS, Mac OS Catalina or later

ii. At least 8GB RAM, and a 256 GB Hard Disk preferably a solid-state drive

5.2.2 Software Requirements

Table 5.1: Software Requirements Table

Software	Details
PyCharm IDE	
Python 3.7.3	• Flask
	• Jinja
	• Pika
	• Pandas
	• NumPy
	• Psycopg2-binary
HyperText Markup Language (HTML) version 5	
Cascading Style Sheets (CSS)	
Java Script	
RabbitMQ / Cloud AQMP Instance	
PostgreSQL Database	

5.2.3: Prediction Module

Data Extraction

The FCT from FAO were in PDF format. Using Adobe Acrobat they were split into three sets, vitamins, minerals and energy, for ease of conversion into CSV. Unnecessary rows were removed by use of a python script and the datasets were then merged into one based on the unique food codes.

Pseudocode to load and validate the food composition table dataset:

```
SET data TO pd.read_csv('FCT.csv')
data.isnull().sum()  # Checking for  null values in the dataset
# Checking info to ensure all values are numeric
data.info()
data.describe()
```

Pseudocode for matrix factorization to retrieve food items ordered by nutritional value:

```
DEFINE FUNCTION predict_food(count, nutrientID):
   SET predictions TO data.nlargest(count, nutrientID)
   RETURN predictions['FoodCode'].tolist()
```

Pseudocode to compute accuracy of prediction function:

```
SET total_accuracy TO 0
SET k TO 10  # number of items to be predicted
FOR deficiency, test_data IN test_dataset.items():
   SET result TO predict_food(k, deficiency)
   SET accuracy TO k - len(list(set(result).difference(test_data)))
   SET total_accuracy TO total_accuracy + accuracy
SET prediction_accuracy TO (total_accuracy/(len(test_dataset)*k) * 100
```

5.2.4 Web Application

5.2.2.1 Patient Facing

Figure 5.1 below is the page the patient sees when visiting the prediction web application, It has collapsible sections to input the biomarkers and nutritional data.

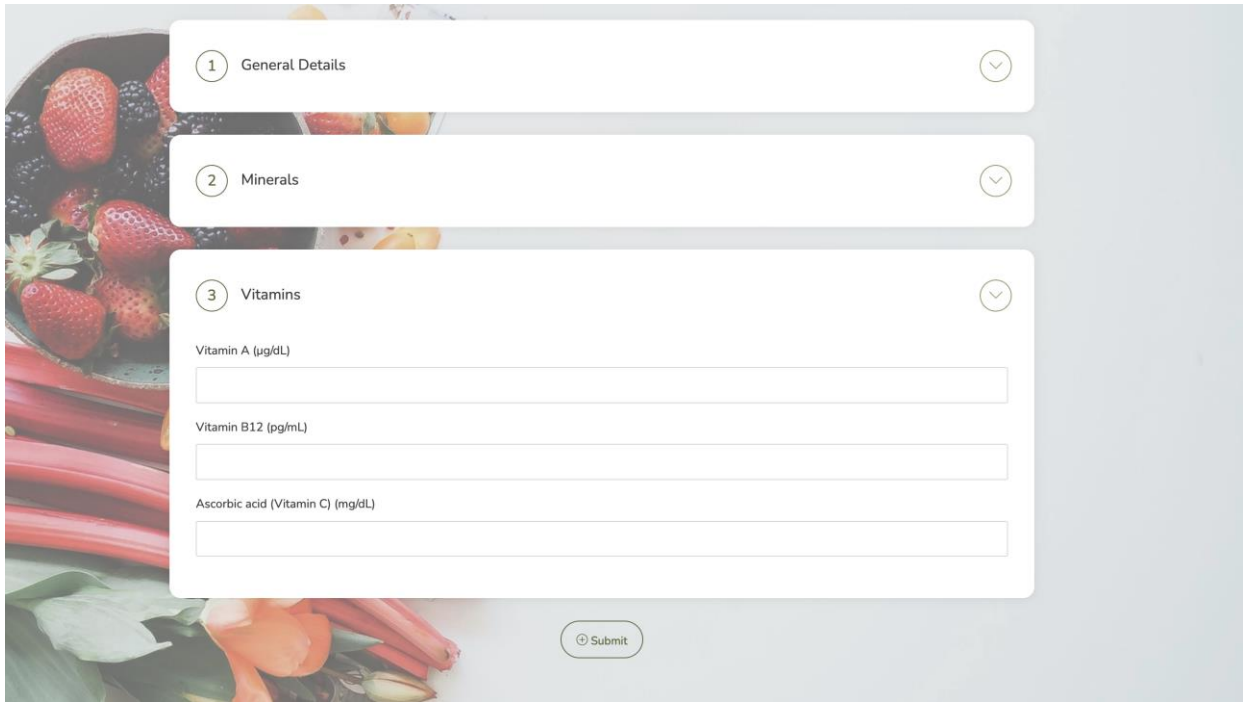


Figure 5.1: User input form

Figure 5.2 is the result page after the patient's input has been processed, it has a BMI section, a list of computed deficiencies and the predictions, both general and location-aware for each deficiency.

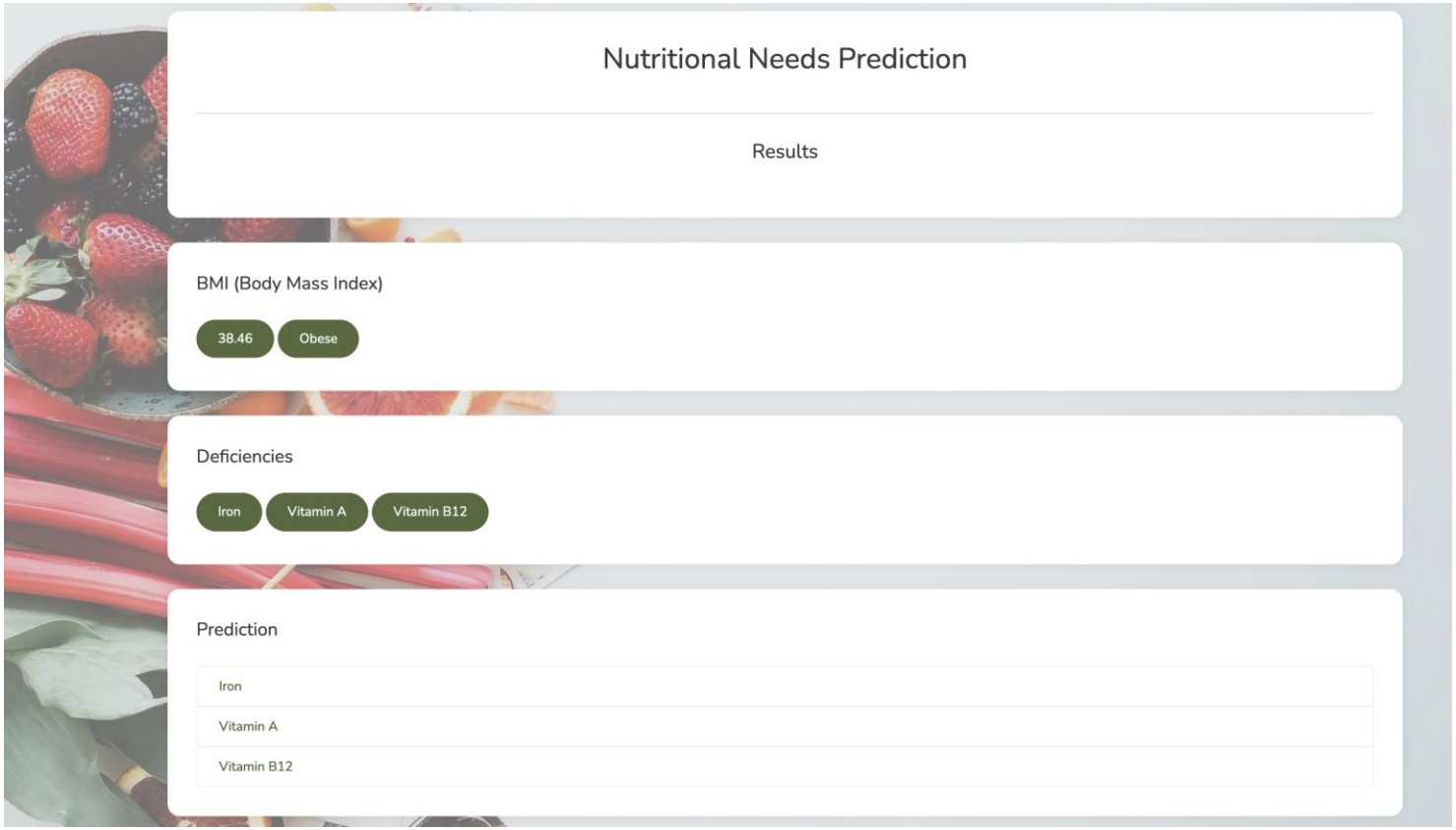


Figure 5.2: Predictions result page

The predictions are collapsible per deficiency as illustrated in Figure 5.3 below and can be individually expanded.

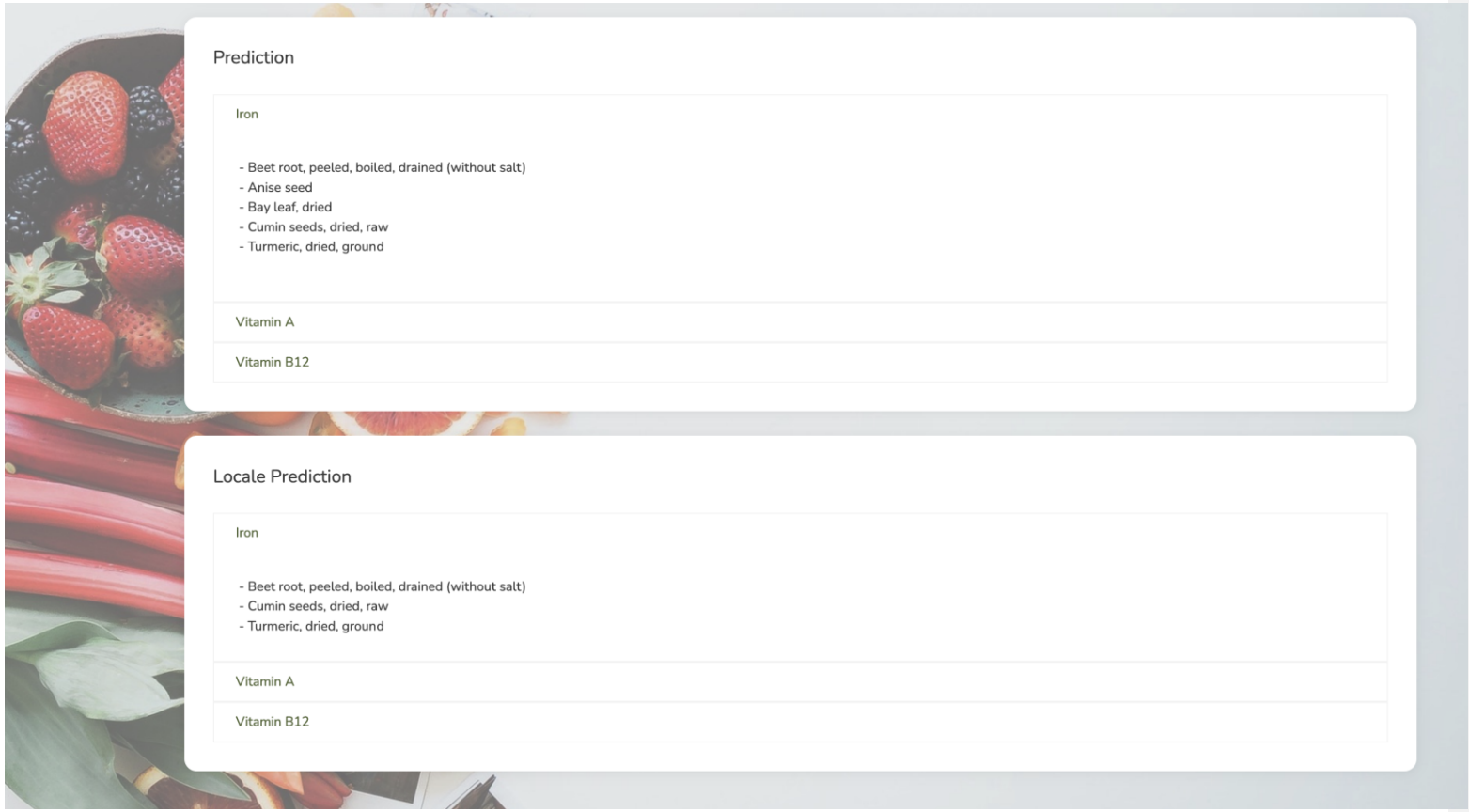


Figure 5.3: Expanded predictions on the result page

5.2.2.2 Nutritionist Facing

Figure 5.4 is the login page to allow nutritionists to access the feedback page.

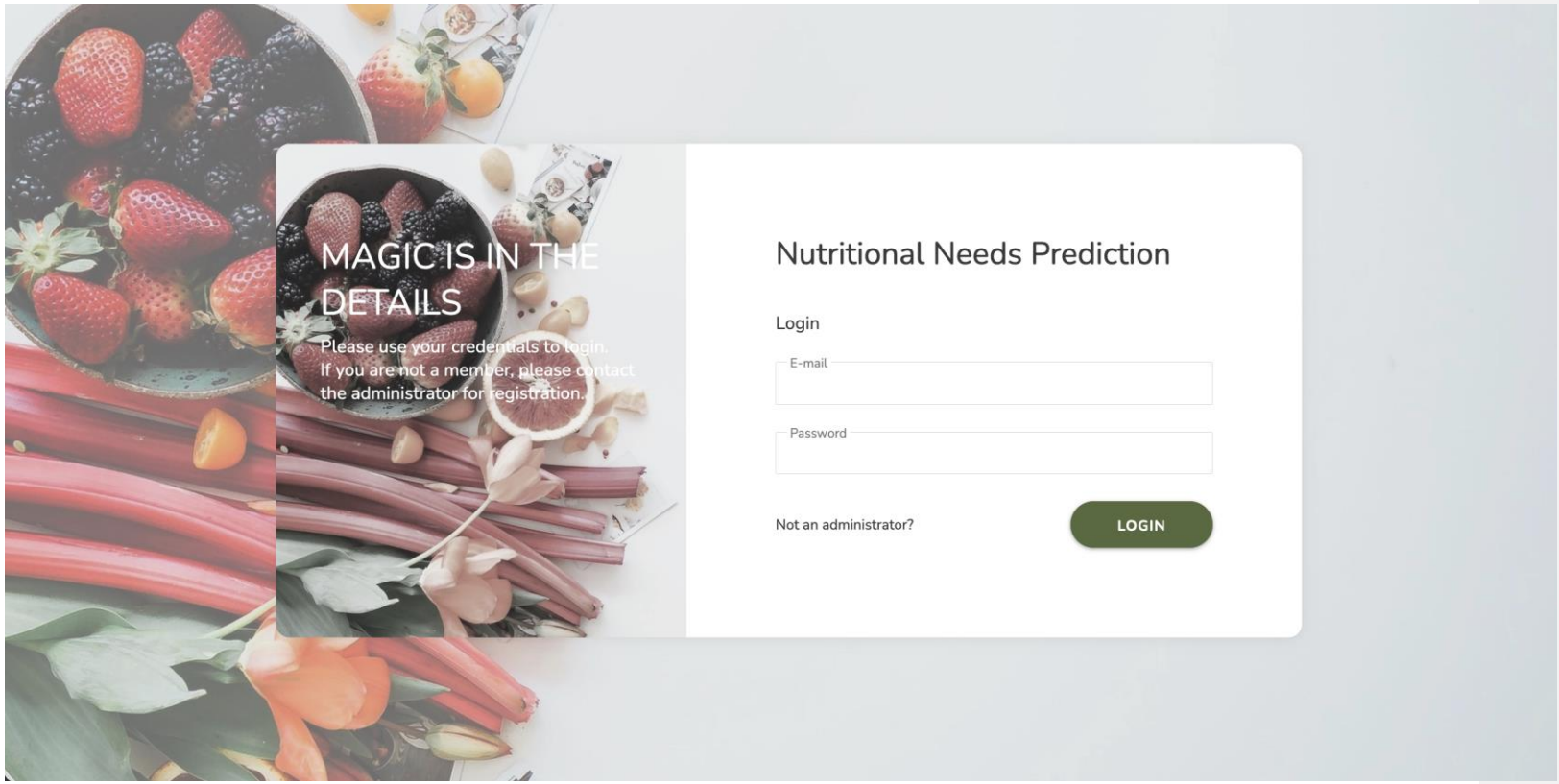


Figure 5.4: Nutritionist login page

Figure 5.5 is the feedback page with instructions on how to mark a prediction as invalid.

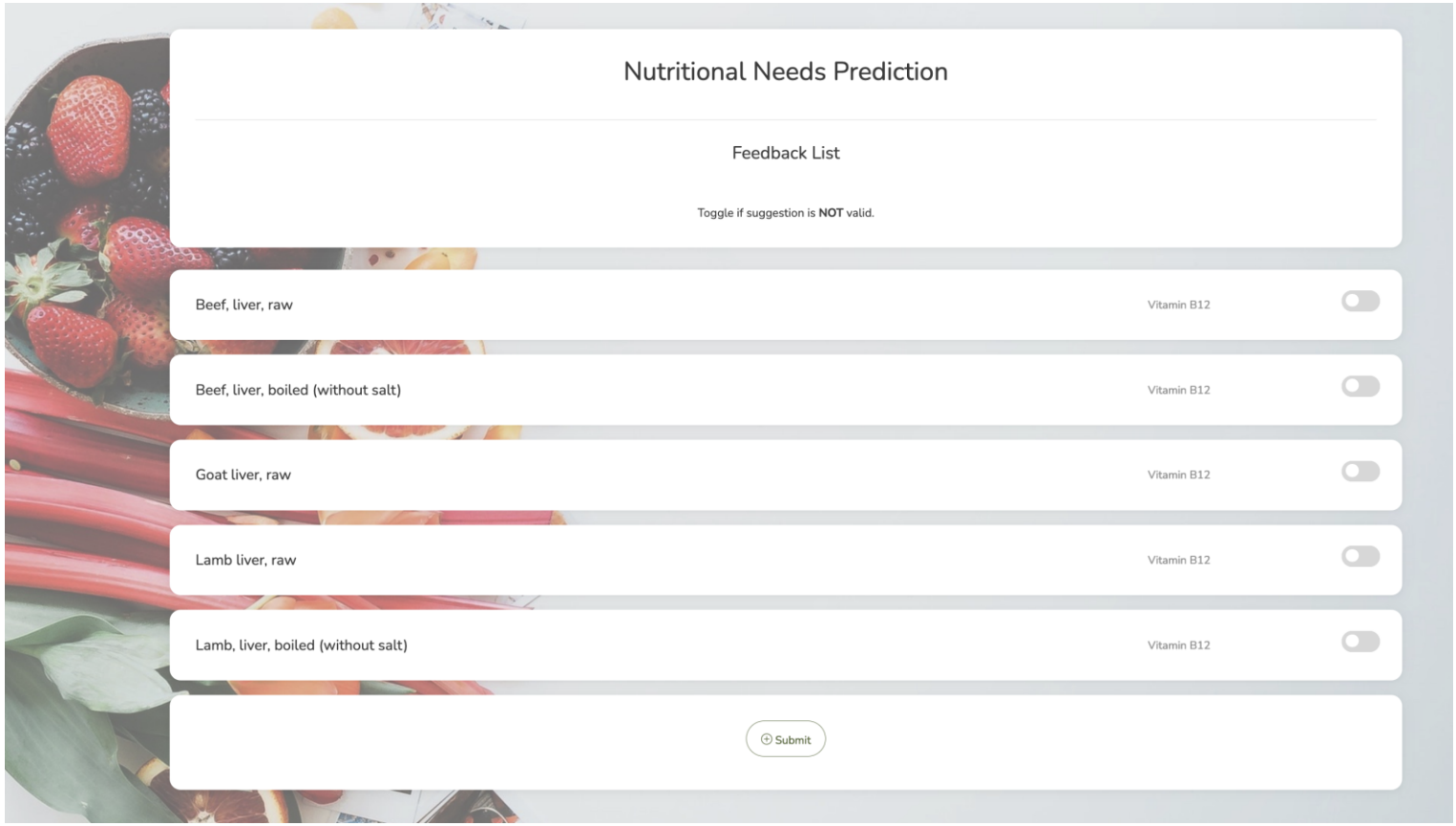


Figure 5.5: Predictions feedback page

Pseudocode to publish predictions to the feedback queue on Cloud AMQP:

```
DEFINE FUNCTION queue_predictions(feedback_lists):
  FOR item IN feedback_lists:
    GET food_name, food_code from DB
    FOR message IN mapping:
      SET message TO food_name + food_code + deficiency
      SET connection TO AMQP_CHANNEL
      SET channel.queue TO ‘feedback’
      channel.publish(message)
      connection.close()
```

Pseudocode to consume predictions from the feedback queue on Cloud AMQP:

```
DEFINE FUNCTION consume():
 SET feedback TO []
 SET connection TO AMQP_CHANNEL
 SET channel.queue TO ‘feedback’
 FOR message IN channel.consume():
      # Acknowledge the message
      IF message is None:
        break
      channel.basic_ack(method_frame.delivery_tag)
      feedback.append(message)
      # Escape out of the loop after 5 messages
      IF method_frame.delivery_tag EQUALS 5:
        break
    SET session['feedback'] TO feedback
```

5.3 System Testing

Some values are mandatory inputs for the user input form, if missing the user is routed to the error page shown in Figure 5.6 which describes the error and an option to return to the user form. This allowed for the testing of input validation. The error page also served to catch any system error instead of displaying a stack trace.

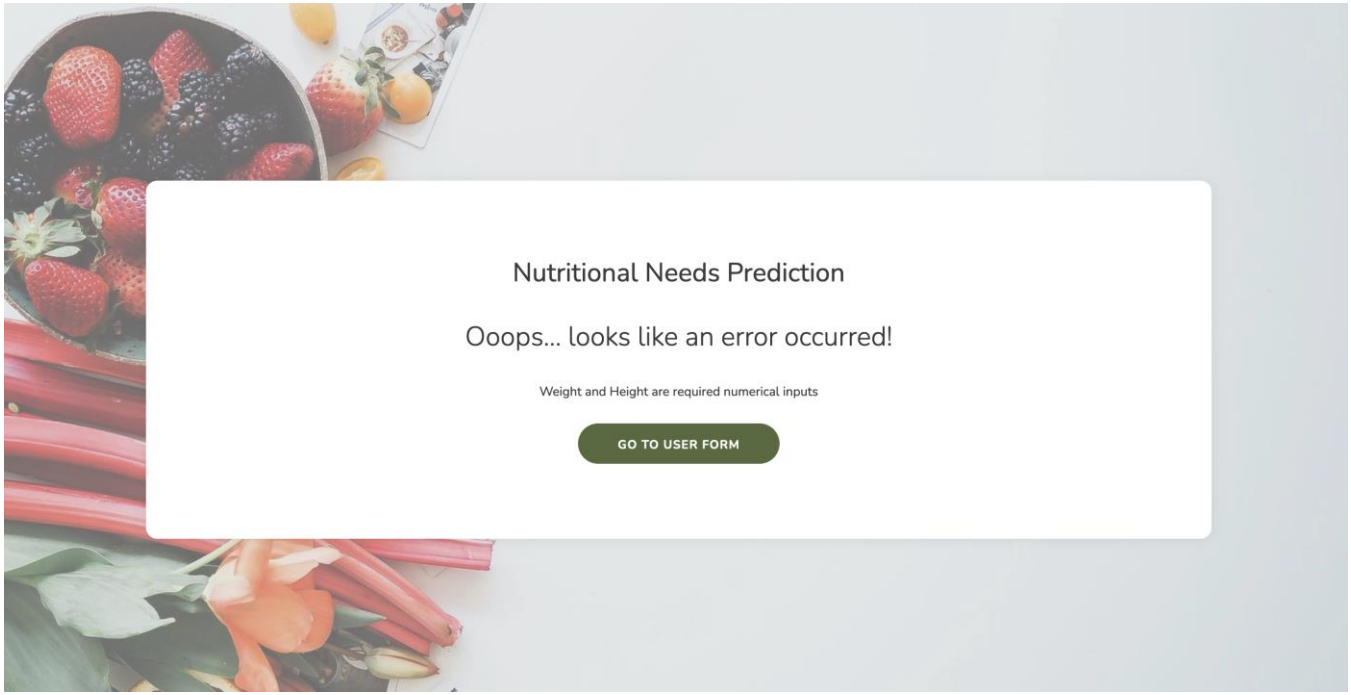


Figure 5.6: System error page

5.4. System Validation

Device compatibility testing was conducted to make sure the web application aligned with the real business environment when deployed. Responsiveness was tested using https://www.responsinator.com/, where one can test screen responsiveness across various mobile and desktop device resolutions.

Table 5.2: Responsiveness Test

Device Name	Orientation	Width	Responsive
iPhone X	Portrait	375px	Yes
iPhone X	Landscape	734px	Yes
Pixel 2 (Android)	Portrait	412px	Yes
Pixel 2 (Android)	Landscape	684px	Yes
iPad (Tablet)	Portrait	768px	Yes
iPad (Tablet)	Landscape	1024px	Yes
MacBook(Laptop)	Landscape	2560px	Yes

5.5 Conclusion

The requirements identified during requirements analysis acted as a guide during the implementation phase. The system design provided guidelines for system development.
The research objectives were the cornerstones that were constantly referred to during the development of the system to ensure they were met.

Commented [VO13]: In chapter 5, move code snippets to the appendix; rather use a brief pseudocode in ch. 5 to explain what the function of the code does

Chapter 6: Discussion

6.1. Introduction

This chapter covers the research conclusions with a specific focus on the research objectives and how they were realized. The study aimed to fill the gap in public awareness by developing a location-aware diabetes nutritional needs prediction tool for Kenya that can be accessed contextually by patients. A web application was developed that used a matrix factorization prediction module after reviewing existing solutions and the approaches behind them.

6.2 Review of Research Objectives

6.2.1 Challenges of providing nutritional awareness to the public

From the literature review on the matter of diabetes specific nutrition awareness, some of the challenges which led to the lack of awareness were identified. The scarcity of dieticians and nutritionists was one key factor (Kenya National Bureau of Statistics, 2019b). These findings are in line with the literature review as discussed in section 2.3. In the reviewed literature, nutritional therapy plays an important role when it comes to diabetes management (Frost et al., 2003; Shafer, n.d.). Visibility of what macro and micronutrients are lacking in our bodies and what foods can help replenish them would go a long way in improving our diets and as a result our health in general.

6.2.2 Evaluation of existing technologies

The second objective was to explore the current technologies used for food suggestions to patients. The research findings were used to help identify the most appropriate technology for nutrition-based, location-aware predictions. The literature review explored different technologies that support food suggestions to patients namely K-Means, random forest, prophet and even touched on Ant colony optimization in one of the existing solutions. The study established that machine learning techniques used in the health informatics sector are playing a key role in tasks such as classification, analysing, association and predictions (Boukenze et al., 2016). The developed application used matrix factorization to implement a location-aware prediction tool.

6.2.3 Development of a nutrition needs prediction tool based on location and nutritional markers

The third objective was to design and develop a prediction tool for type II diabetic patients. The system would need to be location-aware and give predictions based on the user's select location, in this case, the location options were the 47 counties of Kenya.

According to the research finding, patients wanted a prediction tool that would help them with the nutritional management of diabetes. The application developed can guide the patients on what to consume based on the deficiencies identified by nutritional testing and additionally help them eliminate reliance on processed foods and supplements.

6.2.3 Testing the system's ability to give locale-specific predictions based on nutritional markers

The final objective was to test the ability of the prototype. The functionality and responsiveness were tested as evidenced in section five, the prototype successfully predicted food items as per Kenya's FCT and then narrowed the predictions down to the selected location using dummy data that was populated in the database. Furthermore, the functionality of the feedback module was also tested to ensure that once a nutritionist marked a certain prediction as invalid, it would not be predicted again in the future.

Chapter 7: Conclusion and Recommendation

7.1 Introduction

This chapter discusses the conclusions, recommendations, and future work. The objectives are reviewed by examining the formulation of responses to the research questions. Recommendations to the users and stakeholders of the system on how to best leverage it. Future work suggests improvements and advancements that could be made to the system.

7.1 Conclusions

This study emphasizes the importance of nutritional therapy in diabetes management. It is important for patients to know the effect of the foods they consume on their nutritional profile. The prototyped prediction tool analyzes the deficiency levels of various nutrients in the body and predicts food items that would counteract that deficiency. The study also utilizes a data set of locally available foods with the provision to break down the availability to county levels. A review of the implementations currently available and the technologies applied in predictive analytics resulted in those based on patient information to build profiles, those that collate healthy recipes and those based on pathological tests but not necessarily location aware. The main objective of the study was to develop a nutrition needs predictive tool for type II diabetes patients based on location and nutritional profiles.

7.2. Recommendations

This study showed that low-rank matrix factorization can be used in predictive analytics, while not as complex as other machine learning algorithms it still achieves the required functionality and further proves that not every data related challenge requires models to solve it.

The dataset in question is specific to Kenya and contained 651 food items with 27 possible nutritional components providing a total of 17,577 data points. Increasing the food items in the data set would increase the range of predictions that can be provided to the user.

The availability of food items from this dataset per county was not available and dummy data was used in place of location-specific data. The researcher recommends using actual location-specific data in future to improve the validity of predictions given by the system.

7.3. Future Work

The researcher saw that the prototype has the potential for advancements in future, mainly a translation module to allow the user to select a different language such as Swahili or even other local dialects for interacting with the application and viewing predictions.

References

Abbott, D. (2014). *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*. Wiley.

ABIM. (2022). *Laboratory Test Reference Ranges−January 2022*. *January*, 1–12.

Agapito, G., Calabrese, B., Guzzi, P. H., Cannataro, M., Simeoni, M., Care, I., Lamprinoudi, T., Fuiano, G., & Pujia, A. (2016). DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion. *International Conference on Wireless and Mobile Computing, Networking and Communications*, *October*. https://doi.org/10.1109/WiMOB.2016.7763190

Agile Methodolgy. (n.d.).

Alpaydin, E. (2010). *Introduction to Machine Learning*. MIT Press.

Aspin, A. (2020). *Pro Power BI Desktop: Self-Service Analytics and Data Visualization for the Power User*. Apress.

Babcock, J. (2016). *Mastering Predictive Analytics with Python*. Packt Publishing.

Benefits of Blood Testing for Nutritional Deficiencies. (n.d.). https://www.myonemedicalsource.com/2020/06/18/nutritional-testing

Blanchard, B. S., & Fabrycky, W. J. (2010). *Systems Engineering and Analysis* (5th ed.). Pearson.

Blanche, M. T., Blanche, M. J. T., Durrheim, K., & Painter, D. (2006). *Research in Practice: Applied Methods for the Social Sciences*. UCT Press.

Boukenze, B., Mousannif, H., & Haqiq, A. (2016). *Predictive Analytics in Healthcare System Using Data Mining Techniques*. 01–09. https://doi.org/10.5121/csit.2016.60501

Chollet, F. (2017). *Deep Learning with Python* (1st ed.). Manning Publications.

Damasceno, A. (2016). Noncommunicable Disease. In *Heart of Africa: Clinical Profile of an Evolving Burden of Heart Disease in Africa*. https://doi.org/10.1002/9781119097136.part5

Dinov, I. D. (2018). *Data Science and Predictive Analytics: Biomedical and Health Applications using R*. Springer International Publishing.

Eckerson, W. W. (2007). Predictive Analytics, Extending the Value of Your Data Warehousing Investment. *TDWI Best Practices Report*, 34.

Faezipour, M., & Ferreira, S. (2011). Applying systems thinking to assess sustainability in healthcare system of systems. *International Journal of System of Systems Engineering*, *2*. https://doi.org/10.1504/IJSSE.2011.043861

Falk, K. (2019). *Practical Recommender Systems*. Manning Publications.

FAO/GOK. (2018). *Government of Kenya Food Composition*. www.kilimo.go.ke/wp-content/.../KENYA-FOOD-COMPOSITION-TABLES-2018.pdf

Frost, G., Dornhorst, A., & Moses, R. (2003). *Nutritional Management of Diabetes Mellitus*.

Gorakala, S. K. (2016). *Building Recommendation Engines*. Packt Publishing.

Guthrie, R. (2003). *Program Design, Coding, and Testing* (H. B. T.-E. of I. S. Bidgoli (Ed.); pp. 529–543). Elsevier. https://doi.org/https://doi.org/10.1016/B0-12-227240-4/00137-4

Hercberg, S., Castetbon, K., Czernichow, S., Malon, A., Méjean, C., Kesse-Guyot, E., Touvier, M., & Galan, P. (2010). The Nutrinet-Sant?? Study: A web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. *BMC Public Health*, *10*, 242. https://doi.org/10.1186/1471-2458-10-242

Hong, C. (2018). Construction of Corpus in Artificial Intelligence Age. *MATEC Web of Conferences*, *175*, 3037. https://doi.org/10.1051/matecconf/201817503037

Hsu, W. H. (2014). *Emerging Methods in Predictive Analytics: Risk Management and Decision-Making: Risk Management and Decision-Making*. IGI Global.

IDF. (2020). *What is Diabetes*. https://idf.org/aboutdiabetes/what-is-diabetes.html

IDF. (2021). International Diabetes Federation: Diabetes Atlas. In *Diabetes Research and Clinical Practice* (10th ed., Vol. 102, Issue 2). https://doi.org/10.1016/j.diabres.2013.10.013

Jones, T. L. E. (2013). Diabetes Mellitus: the increasing burden of disease in Kenya. *South Sudan Medical Journal*, *6*(3), 60–64. https://doi.org/10.4314/ssmj.v6i3

Kenya National Bureau of Statistics. (2019a). *Kenya population and housing census volume 1: Population by County and sub-County: Vol. I* (Issue November). https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county

Kenya National Bureau of Statistics. (2019b). *Kenya population and housing census volume 1: Population by Socio-Economic Characteristics*.

Knox, S. W. (2018). *Machine Learning: a Concise Introduction*. Wiley.

Leedy, P. D., Newby, T. J., & Ertmer, P. A. (1997). *Practical Research: Planning and Design*. Merrill.

Nithya, B., & Ilango, V. (2017). Predictive analytics in health care using machine learning tools and techniques. *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, 492–499. https://doi.org/10.1109/ICCONS.2017.8250771

O'Hara, E. G., Nuche-Berenguer, B., Kirui, N. K., Cheng, S. Y., Chege, P. M., Buckwalter, V., Laktabai, J., & Pastakia, S. D. (2016). Diabetes in rural Africa: what can Kenya show us? *The Lancet. Diabetes & Endocrinology*, *4*(10), 807–809. https://doi.org/10.1016/S2213-8587(16)30086-9

Palat, J., & Hastie, S. (2021, February 25). *Agile Development Applied to Machine Learning Projects*. https://www.infoq.com/articles/machine-learning-agile/

Pandey, A. K., Rathore, P. S., & Balamurugan, S. (2019). *A Practical Approach for Machine Learning and Deep Learning Algorithms: Tools and Techniques Using MATLAB and Python*. BPB Publications.

Prabhu, C. S. R., Chivukula, A. S., Mogadala, A., Ghosh, R., & Livingston, L. M. J. (2019). *Big Data Analytics: Systems, Algorithms, Applications*. Springer Singapore.

Predictive Analytics Models and Algorithms. (2022). https://insightsoftware.com/blog/top-5-predictive-analytics-models-and-algorithms/

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., & Yang, G. Z. (2017). Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, *21*(1), 4–21. https://doi.org/10.1109/JBHI.2016.2636665

Rehman, F., Khalid, O., Haq, N. U., Khan, A. U. R., Bilal, K., & Madani, S. A. (2017). Diet-right: A smart food recommendation system. *KSII Transactions on Internet and Information Systems*, *11*(6), 2910–2925. https://doi.org/10.3837/tiis.2017.06.006

Research Population. (2009). Explorable.Com. https://explorable.com/research-population

Shafer, S. (n.d.). *Nutrition and Exercise Interventions for Diabetes : Diabetes Type 2 Complete Food Management Program*. 1–33.

Shannon, G. D., Haghparast-Bidgoli, H., Chelagat, W., Kibachio, J., & Skordis-Worrall, J. (2019). Innovating to increase access to diabetes care in Kenya: an evaluation of Novo Nordisk's base of the pyramid project. *Global Health Action*, *12*(1). https://doi.org/10.1080/16549716.2019.1605704

Shmueli, G., & Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, *35*(3), 553–572. https://doi.org/10.2307/23042796

Teng, A. K., & Wilcox, A. B. (2020). A Review of Predictive Analytics Solutions for Sepsis Patients. *Applied Clinical Informatics*, *11*(3), 387–398. https://doi.org/10.1055/s-0040-1710525

Wang, W., Duan, L. Y., Jiang, H., Jing, P., Song, X., & Nie, L. (2021). Market2Dish: Health-aware Food Recommendation. *ACM Transactions on Multimedia Computing,*

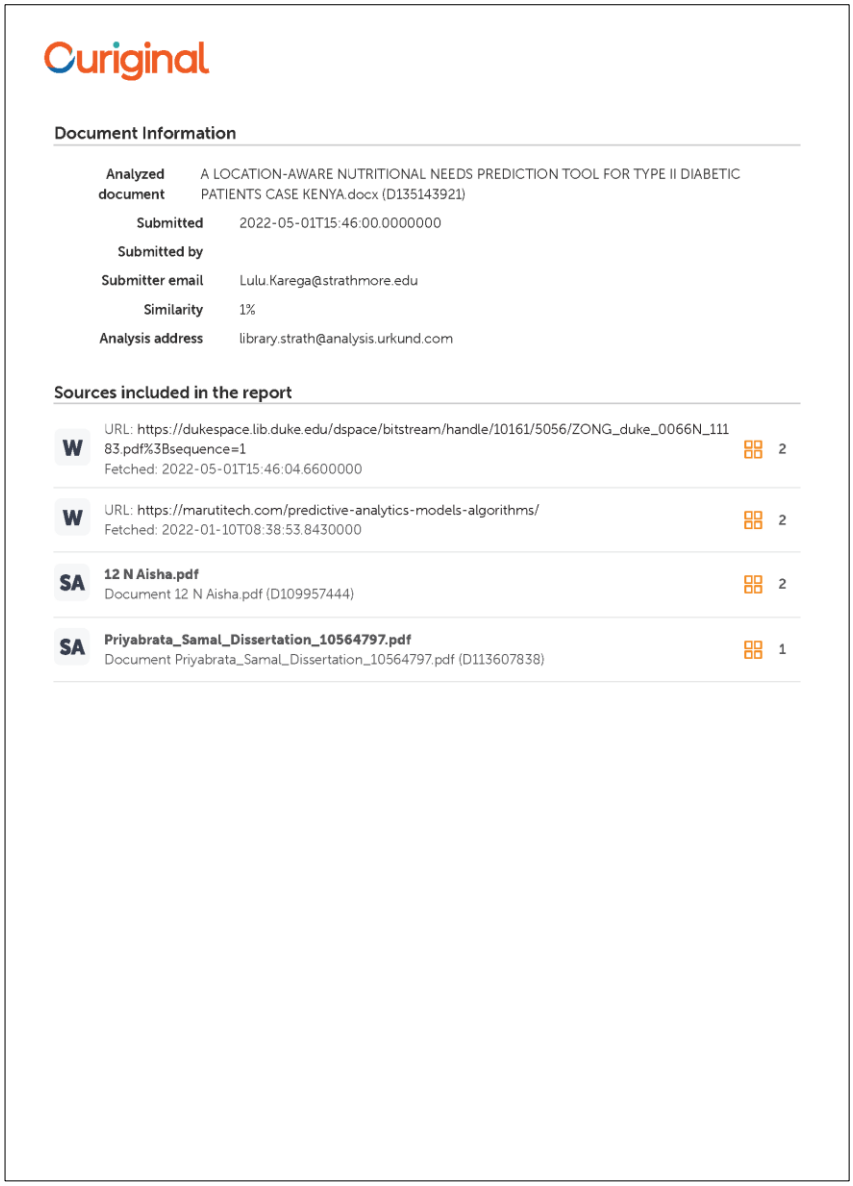
Communications and Applications, *17*(1). https://doi.org/10.1145/3418211

WHO | Kenya faces rising burden of diabetes. (2014, November). https://www.who.int/features/2014/kenya-rising-diabetes/en/

World Health Organization. (2010). Global status report on noncommunicable diseases. In *World Health Organization* (Vol. 53, Issue 9). https://doi.org/10.1017/CBO9781107415324.004

Appendices

Appendix A: Similarity Report



Curiginal

Document Information

Analyzed document	A LOCATION-AWARE NUTRITIONAL NEEDS PREDICTION TOOL FOR TYPE II DIABETIC PATIENTS CASE KENYA.docx (D135143921)
Submitted	2022-05-01T15:46:00.0000000
Submitted by	
Submitter email	Lulu.Karega@strathmore.edu
Similarity	1%
Analysis address	library.strath@analysis.urkund.com

Sources included in the report

W	URL: https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/5056/ZONG_duke_0066N_11183.pdf%3Bsequence=1 Fetched: 2022-05-01T15:46:04.6600000	2
W	URL: https://marutitech.com/predictive-analytics-models-algorithms/ Fetched: 2022-01-10T08:38:53.8430000	2
SA	**12 N Aisha.pdf** Document 12 N Aisha.pdf (D109957444)	2
SA	**Priyabrata_Samal_Dissertation_10564797.pdf** Document Priyabrata_Samal_Dissertation_10564797.pdf (D113607838)	1

Appendix B: Ethical Clearance Confirmation

Strathmore UNIVERSITY

14th March 2022

Ms Karega, Lulu
lulu.karega@strathmore.edu

Dear Ms Karega,

RE: An Index-Based Diabetes Nutritional Educational Corpus for Kenya

This is to inform you that SU-IERC has reviewed and **approved** your above **SU masters'** research proposal. Your application reference number is **SU-IERC1231/21.** The approval period is **11th March 2022 to 10th March 2023.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-IERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-IERC within 48 hours of notification
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-IERC within 48 hours
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to SU-IERC.

Prior to commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

for: **Dr Ben Ngoye,**
Secretary; SU-IERC

Cc: Prof Fred Were,
Chairperson; SU-IERC

STRATHMORE UNIVERSITY INSTITUTIONAL ETHICS REVIEW COMMITTEE (SU-IERC)
14 Mar 2022
Email:ethicsreview@strathmore.edu
P.O BOX 59857-00200
NAIROBI-KENYA

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Appendix C: Raw Data Tables

Table C.1:Distribution of Population Age 15 years and above who owned and used Selected ICT Equipment and Services by age

Age	Population age>=15	Mobile Phone Ownership	Use of Internet	Use of PCs	Use of Online Services
All	28,728,632	20,248,504	9,396,864	4,013,399	1,249,132
15 - 24	9,649,075	4,698,732	3,086,312	1,453,866	320,772
25 - 34	7,333,397	6,151,257	3,254,847	1,336,014	516,006
35 - 44	4,850,354	4,123,888	1,713,260	670,747	247,398
45 - 54	3,062,707	2,587,174	844,963	355,906	112,975
55 - 64	1,973,078	1,571,257	360,461	149,289	41,092
65 - 74	1,165,804	798,353	108,052	38,070	9,137
75+	694,217	317,843	28,969	9,507	1,752

Note: Adapted from Kenya population and housing census volume 1V: Population by Socio-Economic Characteristics by Kenya National Bureau of Statistics, 2019, p 449

Table C.2: Distribution of Population Age 3 years and above who owned and used Selected ICT Equipment and Services by region

Region	Population age >=3	Mobile Phone Ownership	Use of Internet	Use of PCs
Nationwide	43,739,906	20,694,315	9,869,962	4,527,254
Rural	30,253,081	12,249,934	4,139,595	1,614,287
Urban	13,486,825	8,444,381	5,730,367	2,912,967

Note: Adapted from Kenya population and housing census volume 1V: Population by Socio-Economic Characteristics by Kenya National Bureau of Statistics, 2019, p 449

Table C.3:Distribution of Population age 15 years and above who Searched and Bought Goods and Services Online by region

Region	Population age>=15	Used online services
Rural	19,002,236	315,571
Urban	9,726,967	933,562
National	28,729,203	1,249,133

Note: Adapted from Kenya population and housing census volume 1V: Population by Socio-Economic Characteristics by Kenya National Bureau of Statistics, 2019, p 449

Table C.4: Mortality due to diabetes attributable to dietary composition and weight

Category	Deaths in millions (2000-2016)
All Risks	2306
Diet	2006
Weight	300
Processed Meats	10
Red Meat	218
Sugar Sweetened Beverages	72
Overweight	697
Obese	1309

Note: Adapted from the 2021 Global Nutrition Report, Country Profile - Kenya

Appendix D: Web Application Code Snippets

```
def clean_data(scope):
    cleaned_rows = []
    with open(f'{files_dir}/{scope}Formatted.csv', 'r') as read_file:
        rows = reader(read_file, quotechar='"')
        for row in rows:
            pattern = re.compile(r"^\d{5}|^\d{4}")
            if pattern.match(row[0]):
                row[0] = str(row[0]).zfill(5)
                row.insert(1, classes.get(row[0][:2]))
                cleaned_rows.append(row)
        return cleaned_rows
```

Figure D.1: Code snippet for cleaning up CSV files

```
[ ] from google.colab import drive
    drive.mount('/content/drive')
```

```
    Mounted at /content/drive
```

```
[ ] # Importing Libraries
    import pandas as pd
    import numpy as np
    import warnings as w
    w.filterwarnings('ignore')
    print("Importing libraries")
```

```
    Importing libraries
```

```
[ ] df=pd.read_csv('/content/drive/My Drive/ModelDatasets/FinalDataset.csv')
```

Figure D.2: Colab notebook code to load cleaned up file

```
[ ]  df.isnull().sum() # Checking null values in the dataset
```

```
[ ]  df.describe() # Statistics about the dataset
```

	Food Code	Vit A RAE(mcg)	Vit A RE(mcg)	Retinol(mcg)
count	651.000000	651.000000	651.000000	651.000000
mean	7435.307220	383.562596	413.852535	330.978049
std	4943.011764	2805.457097	2830.969063	2770.327872
min	1001.000000	0.000000	0.000000	0.000000
25%	3068.500000	0.000000	0.000000	0.000000
50%	6022.000000	8.000000	10.000000	0.000000
75%	13011.500000	47.000000	60.000000	10.000000
max	15137.000000	38100.000000	38100.000000	38100.000000

Figure D.3: Colab notebook code to validate data

```
[ ]  # Pandas implementation of matrix factorization
     # Predict food items based on the highest nutrient
     # composition with k being the number of items returned
     def predict_food(k,nutrientID):
         result=data.nlargest(k,nutrientID)
         return result['Food_Code'].tolist()
```

```
[ ]
     predict_food(10,'Ca(mg)')
```

```
[13002, 8002, 14004, 10014, 6017, 8018, 13008, 6018, 15101, 1028]
```

Figure D.4: Colab notebook code to implement and test matrix factorization

```
[ ] test_dataset = {
        "Ca(mg)": [13002, 8002, 14004, 10014, 6017, 8018, 13008, 6018, 15101, 1028],
        "Vit C(mg)": [5022,4003,5011,4009,4020,4019,4056,4057,5014,4010],
        "Protein(g)": [8008,8002,15096,3014,15073,7034,7036,6017,3013,7053]
    }
```

```
# Calculating the accuracy of the method on test data
# This was done by use of a formula
# accuracy% = (correct_predictions/total predictions) * 100
total_accuracy = 0
k = 10 # number of items to be predicted
for deficiency, test_data in test_dataset.items():
  result = predict_food(k, deficiency)
  accuracy = k - len(list(set(result).difference(test_data)))
  total_accuracy = total_accuracy + accuracy
prediction_accuracy = (total_accuracy/(len(test_dataset)*k) * 100)
print(f'Prediction Accuracy = {prediction_accuracy}%')
```

```
Prediction Accuracy = 100.0%
```

Figure D.5: Colab notebook code for computing accuracy based on a test dataset

```python

@app.route("/status", methods=["POST"])
def form_post():
    try:
        weight = request.form["weight"]
        height = request.form["height"]
        county = request.form["county"]

        assert weight.replace('.', '').isnumeric()
        assert height.replace('.', '').isnumeric()
        form_data = {k: v for k, v in request.form.items() if v}
        Processor.analyse_responses(float(height), float(weight), form_data, county)
        return redirect(url_for('.results'))
    except AssertionError:
        session['error_message'] = "Weight and Height are required numerical inputs"
        return redirect(url_for('.error'))
    except Exception as e:
        session['error_message'] = str(e).capitalize()
        return redirect(url_for('.error'))


```

Figure D.6: Input validation code snippet

```python
class Predictor:
    def get_prediction_codes(self, deficiencies):
        codes = []
        feedback_queue = []
        types = ["Minerals", "Vitamins"]
        data_frame = pd.read_csv(f'{base_dir}/FinalDataset.csv')
        for deficiency in deficiencies:
            for nutrient_type in types:
                if deficiency in reference_ranges[nutrient_type].keys():
                    column_id = reference_ranges[nutrient_type][deficiency]["dfName"]
                    common_name = reference_ranges[nutrient_type][deficiency]["commonName"]
                    food_codes = data_frame.nlargest(5, [column_id]).FoodCode.tolist()
                    valid_codes = self.get_valid_prediction_codes(column_id, food_codes)
                    print(valid_codes)
                    codes.append({f'{common_name};{deficiency}': valid_codes})
                    feedback_queue.append([valid_codes, deficiency, common_name])
        session['queue'] = feedback_queue
        return codes

    def map_prediction_codes(self, prediction_codes):
        predictions = []
        for prediction in prediction_codes:
            if list(prediction.values())[0]:
                query = self.generate_food_mapping_query(list(prediction.values())[0])
                deficiency_predictions = DB().run_query(query)
                predictions.append({list(prediction.keys())[0]: deficiency_predictions})
        return predictions
```

Figure D.7: Snippet of prediction code

```
<div class="card mb-4">
    <div class="card-body">
        <h5 class="mb-4">Prediction</h5>
        <div id="accordion">
            {% for prediction in session.get('general_predictions') %}
            {% for key, value in prediction.items() %}
            {% set list1 = key.split(';') %}
            <div class="border">
                <button class="btn btn-link" data-toggle="collapse" data-target="#{{ list1[1] }}"
                        aria-expanded="true" aria-controls="{{ list1[1] }}">
                    {{ list1[0] }}
                </button>

                <div id="{{ list1[1] }}" class="collapse" data-parent="#accordion">
                    <div class="p-4">
                        {% for food_name in value %}
                        - {{ food_name[0] }}
                        <br/>
                        {% endfor %}
                    </div>
                </div>
                <br/>
            </div>
            {% endfor %}
            {% endfor %}
        </div>
    </div>
</div>
```

Figure D.8: Snippet of HTML code to render predictions

```python
import pika

from .init_db import DB
from .settings import AMQP_PARAMS as AMQP


def publish(message):
    url = f'amqp://{AMQP["VHOST"]}:{AMQP["PASSWORD"]}@{AMQP["HOST"]}/{AMQP["VHOST"]}'
    params = pika.URLParameters(url)

    connection = pika.BlockingConnection(params)
    channel = connection.channel()
    channel.queue_declare(queue='feedback', durable=True)

    channel.basic_publish(exchange='', routing_key='feedback', body=message)
    print(f'[x] Sent {message}')
    connection.close()


def queue_predictions(feedback_lists):
    for feedback_list in feedback_lists:
        query = "SELECT food_name, food_code FROM food_references where food_code in ("
        for code in feedback_list[0]:
            query = f'{query} {code},'
        query = f'{query[:-1]});'
        mapping = DB().run_query(query)
        for message in mapping:
            to_publish = f'{feedback_list[2]};{message[0]};{message[1]}_{feedback_list[1]}'
            publish(to_publish)
```

Figure D.9: Code snippet to queue predictions for feedback

```python
@staticmethod
def process_feedback(validity):
    types = ["Minerals", "Vitamins"]
    params = validity.split('_')

    for nutrient_type in types:
        if params[1] in reference_ranges[nutrient_type].keys():
            nutrient = reference_ranges[nutrient_type][params[1]]["dfName"]
            query = f'''UPDATE recommendations_feedback SET valid_recommendation = false WHERE
            nutrient = '{nutrient}' AND food_code = {params[0]};'''
            DB().run_query(query)
```

Figure D.10: Code snippet to update predictions validity