Design of meal intake prediction for gestational diabetes mellitus using genetic algorithm

A Project Report submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

Batch no: CSEBNUM_F14

Aishwarya Sharma	**(221810306002)**
Ghanthey Akash	**(221810306019)**
Javvaji Vinitha	**(221810306022)**
S N Srinivasa Yuva Kanishk Rowthu	**(221810306052)**

Under the esteemed guidance of

B.Kumar Babu

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

GITAM

(Deemed to be University)

HYDERABAD

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GITAM INSTITUTE OF TECHNOLOGY

GITAM

(Deemed to be University)

CERTIFICATE

This is to certify that the project report entitled "**Design of meal intake prediction for gestational diabetics mellitus using genetic algorithm**" is a bonafide record of work carried out by **(Batch no:CSEBNUM_F14) : Aishwarya Sharma - (221810306002) ,Ghanthey Akash - (221810306019), Javvaji Vinitha - (221810306022), S N Srinivasa Yuva Kanishk Rowthu – (221810306052)** submitted as part of the requirements for the award of a Bachelor of Technology in Computer Science and Engineering.

Mr.B.Kumar Babu
Project Guide
Department of CSE

Prof. S. Phani Kumar
Professor and HOD
Department of CSE

ACKNOWLEDGEMENT

This project would not have been possible without the assistance of my team members I'd like to thank the individuals who contributed to the project in various ways, as well as those who provided us with exceptional support from the seminar's inception. I am grateful to our honourable Pro-Vice Chancellor, Prof. Siva Prasad, for providing the necessary infrastructure and resources to ensure the success of this seminar.

I am grateful to our beloved Prof.S.Phani Kumar, Head of the Department of Computer Science & Engineering, for giving me the opportunity to work on this project and for her encouragement in finishing this seminar.

I would like to express my heartfelt appreciation to Mr.B.Kumar Babu, Project Guide, Department of Computer Science and Engineering, for their esteemed guidance, moral support, and invaluable advice in ensuring the success of this internship.

I am also grateful to all of the Computer Science and Engineering department staff members who helped make this project a success. I'd like to thank my parents and friends for their assistance, encouragement, and moral support in this seminar, whether directly or indirectly.

Sincerely,

Batch no: CSEBNUM_F14

Aishwarya Sharma (221810306002)

Ghanthey Akash (221810306019)

Javvaji Vinitha (221810306022)

S N Srinivasa Yuva Kanishk Rowthu (221810306052)

TABLE OF CONTENTS:

1.ABSTRACT

The goal of the project is to use algorithms to accurately predict GDM. GDM is also known as pregnancy glucose racism. GDM patients currently use the standard technique to record glucose levels in the blood & keep track of their consuming food. When compared to current cellular health monitoring technology, this approach is ineffective and glucose monitoring is unrealistic in patients with GDM, as discovered when we compared our findings to current cellular health monitoring techniques. Our study's ultimate goal is to create an app where patients & doctors can track their GDM levels.

Despite the fact that many diabetes-related applications have been created, we have yet to find one that is suitable for GDM monitoring. We'll go over how we came up with the idea for the game and how we built it for cell software for using a genetic test to monitor GDM set of rules that seeks to forecast meal intake recommendations intake in this look at.

We will develop a cell application that will allow patients with GDM should keep track of their blood glucose levels while eating. We will investigate the elements of the cell app & discover that the algorithm for prediction correctly expected the outcome on patient's next meal consumption based on glucose levels in the blood. We anticipate that this research will stimulate study into improving self-monitoring apps to help with blood glucose management in GDM patients.

2.INTRODUCTION

Diabetes is a chronic disease characterised by means of a dramatic increase in glucose levels in the blood within a human body. It will be fatal if the patient fails to control and overcome the sickness which will further result in many malfunctioning organs, which will lead to a greater number of chronic illnesses. This severe sickness has the potential to be fatal & can be attributed to a variety of diabetes types. These personalities can be found in both adults and adolescents with varying health histories and genetics. Type 1 diabetes (T1D), type 2 diabetes (T2D), gestational diabetes mellitus, reduced glucose tolerance, and some unusual forms of diabetes, according to Aumiller and Dollahite.GDM is one of the most prevalent and dangerous complications of pregnancy. One of the most frequent pregnancy illnesses is GDM, especially among women who are pregnant. Diabetes mellitus is a disease that affects the body's ability to regulate blood sugar which is unusual and complicated disease that necessitates patients taking excellent care of themselves by preserving consistent medical attention for glycemic control. GDM occurs when the body experiences you may develop a carbohydrate or glucose intolerance, during pregnancy. GDM is a one-time event which is discovered in women who is pregnant during a routine screening. It is only formed during pregnancy and will be released afterward. A GDM offspring, on the other hand, will have an increased possibility of developing type 2 diabetes in the future. This sickness will have an impact on both the mother and the infant that has just been born. Among the most important consequences is that their nutritional development will be harmed.

Pregnant women will have to take special precautions & to have a health examination to determine their glycemic level. Women who have previously suffered from GDM are at a higher risk of having it again in a pregnancy in the future. As a result, they really are advised to undergo screening women with other risk factors are diagnosed early, & if they are normal, to repeat it between 24 and 28 weeks. The glycemic index is a measure of the amount of glucose in the blood in the human body. Basically, the amount of sugar (also known as glucose) in the bloodstream is caused by everything we consume. The pancreatic (human) produces a hormone called 'insulin,' which transports blood to the body for advantage of energy. Nonetheless, diabetic patients are experiencing low insulin levels and are unable to function. As a result, the extra

blood will remain in the circulatory system rather than disintegrating the tissues. GDM patients must maintain a specific level of glycemic control. However, each patient has their own specialised level based on a variety of factors . Age, BMI, health state, and the environment are all aspects to consider and the aftereffect on glycemic control . Consuming foods with a low glycemic index (GI) is one way to lower blood glucose levels. Because of the minor glucose arrival, these foods will reduce glucose response. A diet with a low GI score can assist the patient in reducing his or her need for insulin. Many GDM patients in the recent years have continued to use the traditional manner of recording blood glucose measurements & tracking consuming food. When compared to today's mobile health monitoring technologies, these practises are inefficient. Furthermore, In hospitals health professionals are still using the traditional manner of monitoring glucose levels in GDM patients, which might be inconvenient. Despite the fact that there is a great deal of variation of the usage of self-managing health mobile for GDM has not been completely examined in how mobile health applications are built and used in healthcare. Such a system for regulating glycemic levels and application for meal intake recommendations, in particular, hasn’t yet discovered. Our research's ultimate goal is to develop a GDM monitoring app for patients and doctors. The data will be stored in a centralised database & will be able to communicate with other applications in healthcare. As an initial step, we must develop a meal intake suggestion prediction algorithm in order to keep blood glucose levels stable. The daily meal intake is an important component in maintaining the glucose level. As a result, the proposed application will help them predict their glucose levels based on their daily meals. As a result, patients will be more health-conscious & will be able to avoid any further negative consequences sooner.

Because of technological advancements and the extensive use of data, the digital world is based on the Big Data Framework. The term "big data" implies, it relates to large amounts of data that must be effectively processed and stored. The intelligent world is a huge source of concern for the growth of big data. In terms of text, photos, audio files, and video files represent large amounts of data. Data storage and processing are major issues that must be addressed as the volume of data increases. Because the data is in such large amount, it is difficult to carry out unless it is organised. The more well-organized the data, the more effectively it can be used.

Big data is defined by its volume, velocity, diversity, value, and authenticity. Volume relates to the size of the information, and it essentially tells us how to deal with high adaptability data sets and high dimensional information bases, as well as their handling requirements. The constant appearance of information streams from which valuable data is obtained is characterised by velocity. The nature of data from various sources is determined by its veracity. The term "variety" refers to how different types of data are delivered. Tables, text, sensor data, audio, video, graphs, and a range of other data formats classed as structured, quasi-structured, semi-structured, and unstructured can all be used as source data. Value is critical for obtaining information of value from disparate data sets that vary greatly. We proposed our model in Spark based on the challenges of big data. A framework associated with tools of huge data to assist us in resolving the ability to analyse data generated in diabetes patients health sector. This work aims to create a framework for analysis for predicting the diabetic complications in patients. The purpose of this research is to determine whether or not a user is diabetic based on the user's diet and blood pleasure data.

Because of technical advancements and the excessive use of data, today's digital world is dependant on 'Big Data Processing'. As the name implies, big data technology is capable of effectively managing large amounts of data. The increase in data volume can be attributed to the intelligent world. Data can be in any digital format, including text, numbers, photos, audio, and video. Data storage and management are key concerns that must be addressed as the volume of data continues to grow. The primary need of data management’s scheme is data organization.
Unless data is organised the analytical system's task grows increasingly challenging to process it as the volume of data expands.

The more well-organized the data, the more effectively it can be used. Data posted on social media, commercial platforms, healthcare data, transactional data, and so on are all instances of sample data in this context. Social media data, for example, can be found on social media sites like Facebook, Twitter, and Instagram. As the term "business" is commonly used, the data used by business platforms is massive. Data in healthcare is used by the industry of medicine to analyse patients' medical records from the past and make decisions. Types and formats of data are inconsistent and non-standard. Volume, velocity, diversity, authenticity, and value are the

five data properties that big data technology is predicated on. One by one, all of the terminologies are explained. The amount of data produced by a firm or any other entity is referred to as volume.

Big data is simply a large amount of data that is available from multiple autonomous and heterogeneous sources and is updated in a fraction of a second. Big data has the following characteristics:

a) Volume: This refers to the data's size.

b) Velocity: This is the rate at which data is processed.

c) Variety: This refers to the various types of data generated. Structured, semi-structured, and unstructured data exist.

d) Value: This refers to the accuracy of the data.

e) Veracity: This refers to the data's trustworthiness.

f) Variability: This term refers to the ever-changing nature of Big Data.

The act of collecting, organising, and analysing enormous amounts of data (also known as bi/g data) in order to find patterns and other useful information is known as big data analytics. It is divided into four types. They are as follows :

a. Prescriptive analytics is important but underutilised.

b. Big data is used in predictive analytics to uncover historical patterns and forecast the future.

c. Diagnostic analytics is a type of data analysis that is used to figure out why something happened.

d. Although descriptive analytics, also known as data mining, is at the bottom of the big data value chain, it can be valuable for identifying patterns that provide insight.

Scalability is nothing more than a rapid increase in size. As data volume grows, the value of various data records decreases. Data volume expands faster than computing resources and CPU speed.

One of the key issues of big data is data complexity. Relationships, hierarchies, and numerous data links must also be connected and correlated, or data will spiral out of control.

In the world of Big Data, speed is crucial. This speed is also important in terms of scalability. The speed with which data is generated is determined by two factors: data access time and efficiency.

Accuracy: This is not a serious problem. With the promise of big data comes a plethora of data sources that are not all verifiable. As a result, the accuracy of the data becomes a serious concern.

Cost: Big Data has opened up a world of possible business benefits, and considerable research and discovery is underway to identify the patterns that matter and the insights that turn into value. It is vital to lower the cost of the solutions used to uncover value in order to ensure a successful Big Data project.

Storage: Large and complex big datasets must be managed with dependability, availability, and ease of access. If storage capacity increases in volume and storage, it can be competitive. As a result, data storage research is required.

Security is concerned with data protection. It is a significant impediment to businesses fully utilising their data. A balance should be struck between data privacy and national security.

Processing: It takes a long time to process a large amount of data. To reduce processing time, more research is required.

All of these data characteristics present a significant challenge to big data analysis. As a result, in order to reach the aim, any big data analysis technique must consider these obstacles. With these issues in mind, this paper seeks to propose a big data analytical system for predicting diabetes mellitus in users. This project's purpose is to demonstrate a big data analytics solution for the healthcare business. The purpose of this project is to determine whether or not the user has diabetes mellitus while dealing with the aforementioned challenges. The study is divided into three primary phases to achieve the research goal: data gathering, data preparation, and attribute selection and prediction.

The data is collected in real time, and the data pre-processing phase aims to remove any undesired data from the dataset. The attribute selection is crucial for reducing the proposed

approach's time consumption, and the prediction is carried out using the Extreme Learning Machine, a dependable machine learning method (ELM). The contributions of this work are listed below. Analyze the diabetic in the patients from the dataset using all possible methods in order to determine whether the patient will be born to diabetics as a result of their eating habits and exercise. Using big data analytics, this research aims to develop a diabetes disease prediction system, which is growing more prevalent. The most important step is attribute selection, which reduces the work's computational and temporal complexity. The attribute selection phase chooses the best attributes from the data and sends them to the machine learning algorithm. The sensitivity, specificity, accuracy, and time consumption of the system are all examined.

Diabetes Mellitus is a metabolic condition that lasts a long time. It has an impact on countries with a low and middle income. It is becoming more prevalent in developing countries such as India. India will become a "Diabetic Capital" by 2030. DM is classified into three types:

DM type 1: body failure that insulin isn't produced;
DM type 2: body failure that insulin isn't produced;
and DM type 3: body failure that does not produce insulin.

Today, the majority of people are affected by diabetes, so by implementing this project people will become more aware of the danger of diabetes and the activities that may be done to prevent it as a result of this research. The database in this project can be predicted using data analysis tools. The healthcare business has accumulated a vast amount of data over the last few years, and value-based treatment in hospitals, as well as the digitization of the world, favours computerised information in visible form over hard copy form. Their symptoms are analysed based on their age, height, gender, and family characteristics.

Diabetes mellitus is a category of disorders in which the body's ability to regulate blood maltose levels is impaired. A high maltose content in the blood is a symptom of this category of metabolic disorders. due to the fact that cells do not respond to insulin The well-known symptoms of increased thirst, weariness, itching-skin infection, and headache are caused by a high blood maltose content. There are three forms of diabetes mellitus. 10% of diabetics are

infected with this sort of illness. As a result, the person will need to inject insulin. The body does not utilise insulin adequately in Type 2 diabetes. I'm occasionally afflicted with a total insulin shortage. Non-insulin-dependent diabetes mellitus, or "grown-up-beginning diabetes," was the prior name for this illness. When a kid lady does not include a prior identification of diabetes increase a much greater than the baseline blood glucose rank, the third most common outward emergence estational diabetes occurs. It's probable that type 2 diabetes mellitus will be the first to be improved. As a result of diabetes income in India in 2011, it was estimated that 61.3 million people aged 20 to 79 years will be in this humanity. The big data idea is used to convey the high volume of data that we are employing in this project.

A journalistic assessment recognises numerous diabetes-related outcomes disseminated through unique diabetes-related methods and resources in India. As a result of the numerous calculation models with datamining systematic technique that may be used to predict diabetes, a huge number of people have been urbanised. To anticipate the arrival of the pre-processed dataset, a standard neural network model is utilised, with the missing importance substituted by the area of the matching feature. The results obtained referring to the level of danger that lies ahead of both your heart bother or knock. The work of fiction pre-processing point with absent worth declaration for mutually statistical and unconditional information. To impute missing continuous values, a fusion of arrangement and failure grass (CART) and Genetic Algorithms was improved, and Self-Organizing Feature Maps (SOFM) to impute categorical values. To support and connect data in a given area, a physical condition information exchange (HIE) warehouse is being deployed. A reliable information distribution system.

This information and electronic communication system distribution allows patients to be admitted to military physical condition while also promoting alternative types of care that are more appropriate for them. It identifies which patients require more attention and care than others. It gives you the knowledge you need to figure out the methods you'll need to follow to get the most out of positive performance adaptation. Operations management, health check management, biomedicine, categorization design, and planning all benefit from projecting analytics Focusing on the cost of patients who are regularly admitted and readmitted to a nursing home for chronic diseases that are identical or multiple is one of the challenges that a healthcare

projecting analytics system may help with. Various big data skill sets are loaded and investigated in relation to health care as a whole, with the help of good organisation and cost investments in better healthcare. In health care, the Hadoop custom has increased in importance in terms of processing data and accepting a wide range of data management tasks. Hadoop-based analytics on mutual computation and storage space can boost cost efficiency in the direction of being present.

Diabetes is now one of the world's fastest growing diseases. Diabetes mellitus, also known as diabetes, is a chronic, metabolic disease characterised by elevated levels of blood glucose (or blood sugar), which leads to serious damage to the heart, blood vessels, eyes, kidneys, and nerves over time, according to the WHO.

Diabetes is classified into three types: type 1, type 2, & gestational diabetes. Juvenile diabetes is another name for type 1 diabetes. Type 1 diabetes develops when the human body's ability to produce insulin fails. Type 2 diabetes develops when the human body does not properly utilise the insulin that it produces. Gestational diabetes is characterised by high blood glucose levels in pregnant women and is associated with complications for both mother and child. GDM normally goes away after pregnancy, although it increases the risk of type 2 diabetes in women and their children later in life.

The following are the contributions of this work.

1.Using big data analytics, this research aims to develop a diabetes illness prediction system, which is incredibly common nowadays.

2.The most important step is attribute selection, which reduces the work's computational and time complexity.

3. The best attributes from the data are selected and sent to the machine learning algorithm in the attribute selection step.

4. The sensitivity, specificity, accuracy, and time consumption of the system are all examined.

3. Literature Survey

3.1 Base Paper

Hypertension is more common in diabetics, with prevalence varies based on diabetes type and duration, age, gender, race/ethnicity, BMI, glycemic management history, and the presence of kidney disease, among other factors. Diabetes hypertension and poor diabetes outcomes have a strong epidemiological relationship. Clinical trials suggest that medication therapy is more successful than placebo at lowering these outcomes and achieving a blood pressure target of 130/80 mmHg. Many people will undoubtedly require three or more medicines to achieve the recommended dosage. The achievement of the target blood pressure goal with a regimen that has minimal side effects and is affordable for the patient is likely more essential than the pharmacological strategy.

It has been well documented in the authors that subjects with peripheral neuropathy caused by diabetes mellitus are at a high risk of developing foot ulceration. Reduced sensation, in conjunction with high underfoot pressures, has been identified as major aetiology factors in the development of plantar naturopathic ulceration. This study looked at how four orthotic treatment techniques affected the pressure on the metatarsal head, heel, and toes while walking. A diabetic patient group underwent a pressure measurement research while wearing four insoles, one of which was built using the computer model proposed in this project. The goal of this study was to see how different insoles prescribed and manufactured using different processes affected plantar pressure in a group of diabetes individuals in the early stages of the condition. Four different types of insoles were designed and manufactured utilising market-available methods; the insole was built and manufactured using the computer model recommended for this project. According to the findings, each patient requires a personalised insole and, in many circumstances, an unique examination.

The designers of A molecular hyperspectral imaging (MHI) system was designed to examine the preventive effects of erythropoietin (EPO) on early diabetic retinopathy in rats. Hyperspectral pictures of rat retinal slices from three separate groups were taken using the system: normal, diabetic, and EPO. As biological parameters, the spectral transmittance index, the thickness of the outer nuclear layer, and the cell area % were all defined. These were also presented, along

with the accompanying procedures for computing them. The findings of the trials demonstrate that after treatment, the EPO group's newly defined biochemical characteristics resemble those of the normal control group more than those of the diabetic group. This implies that EPO has some protective effects on the retinal cells of chemically induced diabetic rats after being injected intravitreally at the outset of diabetes. The findings also suggest that the MHI approach could give useful quantitative information about retinal sections for ophthalmologists to employ in determining the pathophysiology of diabetic retinopathy and assessing the preventive effect of EPO on the disease.

According to the authors It has been widely stated that classical ontology is insufficient for dealing with imprecise and ambiguous knowledge in some real-world applications such as personal diabetic diet recommendation. Fuzzy ontology, on the other hand, can effectively aid in the handling and processing of uncertain data and knowledge. This paper proposes type-2 fuzzy ontology (T2FO), a novel ontology model based on interval type-2 fuzzy sets (T2FSs), with applications to knowledge representation in the field of personal diabetic-diet recommendation. The T2FO is made up of three parts: 1) a type-2 fuzzy personal profile ontology (FPPO); 2) a type-2 fuzzy food ontology (FFO); and 3) a type-2 fuzzy-personal food ontology (FPFO) (type-2 FPFO). Furthermore, the paper introduces a T2FS-based intelligent diet-recommendation agent (IDRA), which includes 1) T2FS construction; 2) a T2FS-based personal ontology filter; 3) a T2FS-based fuzzy inference mechanism; 4) a T2FS-based diet-planning mechanism; 5) a T2FS-based menu-recommendation mechanism; and 6) a T2FS-based semantic-description mechanism. In the proposed method, domain experts first plan the diet goal for the involved diabetes and create nutrition facts for common Taiwanese foods. Second, the diabetics involved are asked to routinely enter eaten items. Third, the ontology-creating mechanism builds a T2FO, which includes a type-2 FPPO, a type-2 FFO, and a set of type-2 FPFOs. Finally, the T2FS-based IDRA retrieves the created T2FO in order to recommend a personalised diabetic meal plan. The experimental results demonstrate that the proposed approach is effective and that the menu can be used as a reference for the diabetes patients after diet validation by domain experts.

According to the author, a model for diabetic diet is a multiobjective optimization problem, as opposed to the usual single-objective method. We create a diabetes nutrition diet decision-

making model in this research. The state transition table is used to characterise the genetic operators. The nutrition decision-making optimization problem for diabetic patients is solved using the multi-objective optimization method NSGA-II. The algorithm can effectively address diabetic patients' nutrition optimization problems and give diet designers with powerful nutrition decision support, according to the findings of the experiments.

Diabetes is one of the metabolic diseases that has been explained in terms of its effects on productivity and human quality of life. Diabetes management is a difficult task for patients because it requires them to monitor and control their blood glucose levels in order to avoid serious diabetic complications. It is also difficult for doctors to manually interpret large amounts of blood glucose data in order to tailor treatment to each patient's needs. Diabetes should be treated with a balanced diet and physical activity, which can help patients avoid serious complications. As a result, there is a need for technology to assist diabetics and doctors in controlling the disease and reducing its complications. Various techniques have been employed for Diabetes diet recommendation. Diabetes diet recommendations have been made using a variety of techniques. However, any technique to deal with the available uncertainties associated with varying people's opinions and preferences is required. This paper describes our ongoing work to create a type 2 fuzzy logic-based diet recommendation system for diabetes to aid in the achievement of a healthy lifestyle in order to control the disease.

According to the scientists, diabetes is a dangerous chronic disease that can be controlled with a well-balanced diet and frequent exercise. Gender, weight, height, age, needed calories and nutrition values, food and exercise preferences, clinical standards, and current vital signs are just a few of the choice variables that go into maintaining a healthy diet and exercising properly. For diabetic patients, we developed a semantic rules and reasoning-based strategy for providing food and activity recommendations. The name of the prototype application is Semantic Healthcare Assistant for Diet and Exercise (SHADE). Individual ontologies, together with SWRL rules, are defined for several domains (person, sickness, diet, and exercise), and then imported into an integrated ontology. The integrated ontology creates suggestions as conclusions based on data and rules using the Pellet reasoner. Each created meal menu is a list of food products with portion amounts so that the menu is customised, healthy, and balanced. Finally, SHADE offers

exercises depending on the user's favourite activities, as well as the duration and intensity of such activities.

The goal of this study is to use free-living data to model the blood glucose metabolism of type 1 diabetic patients. The proposed method takes into account the influence of diet, medication, and exercise on blood glucose levels. Compartmental models are used to quantify subcutaneously administered insulin absorption, glucose absorption from the gut following a meal, and the effects of exercise on plasma glucose and insulin dynamics. To estimate subcutaneous glucose concentrations, compartmental analysis is combined with a glucose predictive model that employs Support Vector Machines for Regression. The model was trained and tested using real-world data from two type 1 diabetic patients. The obtained results demonstrate the model's ability to predict glucose response with sufficient accuracy.

According to the authors' scenario, the desire for a healthier diet is growing by the day, particularly among diabetics, where the goal of a healthier diet can only be achieved by keeping track of daily food intake and glucose levels. As a result, there is an ever-increasing demand for automatic tools that can assist diabetics in managing their diet and also assist physicians in better analysing the effects of various types of food on diabetics' glucose levels. In this paper, we propose an intelligent food recognition and tracking system for diabetics, which could be an important component of a mobile application that we propose to use glucose measuring sensors to couple food intake with blood glucose levels. For food recognition, which is a critical component of the application, we rely on several feature extraction and classification techniques, which are individually and jointly used using an early and two different late fusion techniques, namely I Particle Swarm Optimization (PSO) based fusion and (iii) simple averaging. Furthermore, we assess the performance of several deep features. . Furthermore, we collect a large-scale dataset containing images from various types of local Middle-Eastern food, with the goal of making it a powerful support tool for future research in the domain.

Diabetes, according to the authors, is a disease that impairs the human body's ability to store and regulate sugar. This disease develops as a result of an excessive intake of sugary foods, excessive work pressure, or unbalanced routines lacking in proper diet. Diabetes, once developed, is more

difficult to overcome and thus has an impact on human body functioning, leading to the failure of many human body parts. One of the most serious complications of diabetes is vision loss, which is technically known as Diabetic Retinopathy (DR). Automated analysis of retinal images is evolving and proving useful in the screening and early detection of diabetic retinopathy. Early detection of DR aids in overcoming the disease and preventing blindness. Many studies have been conducted in recent years to develop automated tools that could help simplify screening of affected images and reduce evaluation time in disease detection. This work aims to investigate the algorithms for the detection and classification of diabetic retinopathy, as well as to present a brief summary of selected relevant work from the last 5 years. A comparative study of the most recent techniques for screening diabetic retinopathy through the analysis and segmentation of lesions is presented.

The authors present a new paradigm for modelling illness in the human population in their paper. In this paper, we propose using a Mobile Software Agent to create a patient model. We focus on Diabetes Mellitus because of the disease's prevalence and the fact that many citizens must learn to manage their disease through simple dietary, exercise, and medication guidelines. This type of illness modelling has the potential to predict outcomes for diabetic patients based on their lifestyle. We also believe that if patients are not meticulous in controlling their blood sugar and insulin levels, the Patient Agent could be an effective tool in assisting them to understand their prognosis. The Patient Agent was created in accordance with the general parameters used in standard Diabetes medical tests. Input variables such as Food, Exercise, and Medications, as well as other risk factors such as Age, Ethnicity, and Gender, have been transformed into output variables such as Blood Glucose and Blood Pressure using conventional formulae. The outputs of the Patient Agent deteriorate over time, representing the Patient Agent's time evolution.

From the writers of Food and nutrition are essential for good health. It is critical for everyone to eat a healthy diet, especially diabetic patients who have several limitations. Nutrition therapy is a major solution for preventing, managing, and controlling diabetes through nutrition management, based on the belief that food provides vital medicine and maintains good health.
Typically, diabetic patients must avoid extra sugar and fat, so the food pyramid is recommended to patients in order to find a substitute from the same food group. However, there is still dietary

diversity within food groups that can have an impact on diabetic patients. In this study, we proposed a Food Recommendation System (FRS) for diabetic patients using food clustering analysis. In the context of nutrition and food characteristics, our system will recommend the appropriate substituted foods. For food clustering analysis, we used Self-Organizing Map (SOM) and K-mean clustering, which is based on the similarity of eight significant nutrients for diabetic patients. Finally, nutritionists evaluated the FRS, finding it to be very effective and useful in the nutrition field.

The main goal of this project, as proposed by the author, was to develop and implement a work-based network environment at Southwest Research Institutereg (SwRIreg) to monitor the arterial health of SwRI employees. The goal of this implementation was to create a network of existing medical devices and investigate how to improve employee health by utilising the developed network. Existing blood pressure monitors were retrofitted and linked to the SwRI network. To collect data measured by blood pressure equipment and present it to users, a Web-based system was developed using open source software components and LabVIEWtrade. The programme enrolled a total of 109 SwRI employees. 1705 blood pressure measurements were taken over a 7(1/2) month period during which blood pressure stations were installed. The average systolic and diastolic measurements were 124 and 74, respectively, with 85 percent of the measurements falling into the normal or prehypertensive range. Overall, 63.3 percent of participants used a single blood pressure station to obtain measurements, usually the one closest to their workplace.
In the coming years, healthcare costs are expected to rise by double digits. Cardiovascular disease is the leading cause of death in men in the United States, accounting for a sizable portion of healthcare spending. Hypertension, if left untreated, can lead to a heart attack or even death. When detected early, hypertension can often be reversed with diet, exercise, and pharmaceuticals, lowering the risk of catastrophic disease, hospitalisation, and death.

However, the authors claim that as a patient's condition worsens, diet management becomes more difficult. To prevent ischemic heart disease, people with metabolic syndrome must engage in both self-management and behavioural changes. We improved the system to support behaviour change in people with metabolic syndrome in this study. The system has three functions: entering the user's lifestyle information, calculating the total amount of calorie intake,

and referring to food model pictures of standard quantities of 80 kcal for behaviour change. As the first step in changing behaviour, we focused on a self-review of the user's lifestyle. We conducted an experiment with the help of 14 medical staff members to test the system's functionality. There were no major issues with the system. We also put the system through its paces with health-check examinees. The results of these tests suggested that our system will assist physicians in advising health checkup examinees on total calorie intake. In 2008, the World Health Organization reported that ischemic heart disease was the leading cause of death worldwide. Obesity, hypertension, arteriosclerosis, and other metabolic abnormalities are the leading causes of ischemic heart disease. These conditions are linked to lifestyle factors like diet and exercise. Using a personal digital assistant, we created a system to assist patients in self-management (PDA).

Hypertension is known as the "silent killer" in the medical community. It is regarded as a risk factor for chronic kidney disease, heart failure, impaired vision, ischemic heart disease, stroke, and other conditions. Systolic and diastolic blood pressure are the two components of hypertension. According to studies, 90-95 percent of the cause of hypertension is a change in lifestyle, so diet is very important for hypertension patients. According to WHO studies, chronic disease-related deaths in India have increased by 18%. However, 1.13 billion people worldwide were affected by high blood pressure. In 2015, the observed systolic blood pressure measurement is greater than 140 mmHg, and the observed diastolic blood pressure measurement is greater than 90 mmHg. Using the Analytic Hierarchy process, the paper demonstrates the best diet plan for hypertension patients. The technique used in this paper to represent a diet plan is unique and has not been demonstrated previously. The diet plan takes into account all of the meals that a hypertension patient must consume for breakfast, lunch, and dinner. The Entropy method is used to validate the results. The results obtained during validation are the same as those obtained using AHP.

According to the author In contrast to the traditional single-objective method, hypertension design is modelled as a multi-objective optimization problem. We present an improved multi-objective genetic algorithm to solve hypertension nutritional diet problems in this paper. Simulated annealing is presented to overcome shortcomings in multi-objective genetic algorithms such as poor local search and premature convergence. The experimental results show

that this algorithm is very effective for hypertension design and provides the designer with powerful decision support.

In Hypertension is becoming a major public health concern around the world. People have a hectic lifestyle and eat unhealthy foods. The rate of Non communicable Diseases (NCDs) such as hypertension, as well as the rate of death caused by such diseases, is increasing as a result of poor eating habits. The paper proposes a DASH diet recommender system that recommends healthy Mauritian diet plans to hypertensive patients in order to promote healthy eating habits in Mauritius. The system includes a recommendation engine that recommends personalised diet plans to hypertensive patients based on factors such as age, user preferences for food, allergies, smoking level, alcohol level, blood pressure level, and dietary intake. The system employs a mobile application that is both convenient and quick to use. According to a survey, the application has assisted users in controlling and lowering their blood pressure.

According to the authors in Studies on health care services have been actively conducted in the recent ageing society in order to provide quality services to medical consumers in wire and wireless environments. However, there are some issues with these health care services, such as a lack of personalised service and a uniformed approach to services. Studies on customised services in medical markets have been conducted to address these issues. However, because a diet recommendation service is only concerned with the user's personal disease information, it is difficult to provide users with specific customised services. In health care services, this study offers a customised diet recommendation service for preventing and managing coronary heart disease. This service offers a customised diet to customers based on their basic information, vital signs, family history of diseases, food preferences based on seasons, and intakes for customers who are concerned about coronary heart disease. Users who receive this service can use a customised diet service that differs from the standard service and that supports continuous services and helps customers change their living habits.

In The small corpus data is built up based on the text data of traditional Chinese medicine diet, taking into account the entity recognition feature of Chinese medicine, and the entity recognition experiment is carried out on the self-built corpus data. In this paper, we train and test the

conditional random field method, which is based on a machine learning model, on common clinical diseases such as colds, coughs, hypertension, hyperlipidemia, and diabetes. The target entity is extracted in clinical manifestations of body parts and dialectical clinical manifestations of two categories. The experimental results show that by adding structural features to sentences and expanding the corpus data, the proposed method improves accuracy and recall rate, and the validity of adding structural features is validated.

From the writers of Cholesterol, hypertension, and diabetes are the three major chronic diseases that the majority of people suffer from, and these people frequently use search engines to learn more about these issues. However, almost all diet-related information on the internet is not suitable for people to gather information about diet suggestions. This paper proposes a diet suggestion system that can advocate a prudent diet for such people. We created a system that recommends a healthy diet based on our knowledge of the three highly chronic diseases mentioned above. Using the optimization algorithm known as unbounded knapsack, we propose a solution to the menu recommending problem. We created a model that meets individuals' nutritional needs while enforcing the "Laws of Nutrition," a set of hypotheses used by nearly all Latin American nutrition scientists. This prototype corresponds to a constraint-based numerical optimization problem. We create a menu items generator application model to create a user-friendly menu with various properties.

COSTAR, a Computer-Stored Ambulatory Medical Record system, has been in use in the Family Practice/Primary Care Clinic at Pease Air Force Base, N.H., since November 1981, according to the authors. This paper describes a computer-based programme that nurses can use to help them manage a hypertension screening and education programme for known and suspected hypertensives. The programme includes the following components: 1) a data record for five-day blood pressure screening and patient education data; 2) a patient-specific instruction sheet containing information about the patient's medications and diet; and 3) a drug interaction check list informing the nurse of potential interactions of drugs active for that patient. The computer-generated instruction sheet has been evaluated by patients.

Traditional time-unspecified single measurements of blood pressure and heart rate may be misleading because they can be influenced by the patient's emotional state, position, and diet, as well as external stimuli. The assessment of predictable variability in blood pressure and heart rate by (1) using fully ambulatory devices and (2) properly processing the resulting time series can be useful in assessing early cardiovascular disease risk in pregnancy. This method was used to investigate the characteristics of blood pressure variability during pregnancy in order to detect blood pressure elevations early.

According to the authors, in recent years, people have become more concerned about eating a healthy diet as a result of advancements in information science and technology, so diet types have gradually changed and have become more focused on health management. In today's Taiwan, people have an irregular life, a long-term unhealthy diet, work pressure, and other factors that contribute to chronic disease, such as diabetes, hypertension, and high cholesterol, among others. However, the majority of dietary recommendation systems are unable to provide dietary recommendations to patients suffering from chronic diseases. Patients are concerned about what foods are edible, but they are unconcerned about whether the nutrients are balanced. As a result, this study developed a chronic disease diet recommendation system based on expert knowledge, and provided chronic diseases with more convenient and precise dietary recommendations. To build the recommendation system in this study, we used ontology, decision trees, and Jena. Dietitians evaluate the dietary recommendations, and the verification accuracy is 100 percent. As a result, this dietary recommendation system can provide nutrient dietary recommendations for chronic disease patients in order to achieve a convenient and healthy diet.

.

Python

Python is a programming language, which means that both humans and machines can understand it. Python was invented by Guido van Rossum, a Dutch software programmer, to address various issues with computer languages of the time.

Python is a high-level computer language for general-purpose programming that is interpreted. Python was created by Guido van Rossum in 1991 and features a design philosophy that emphasises code readability as well as a syntax that allows programmers to express concepts in fewer lines of code with significant whitespace. It contains techniques that enable clear programming on small and big scales.

Python has a dynamic type system as well as automatic memory management. It supports a wide range of programming paradigms, including object-oriented, imperative, functional, and procedural, and it includes a large and comprehensive standard library.

For a wide range of operating systems, Python interpreters are available. C Python, the reference Python version, and nearly all of its variant implementations, are open source software with a community-based development approach. The Python Software Foundation, a non-profit organisation, is in charge of C Python.

You Can Use Python for Pretty Much Anything

Python is a general-purpose language that can be used in a wide range of projects, which is a significant advantage of learning it. The following are just a few of the most common fields where Python has been used:

Data science is the study of data. Scientific and mathematical computing is the study of numbers.

Web development Computer graphics Fundamental game development Mapping and geography (GIS software)

Python Is Widely Used in Data Science

Python's ecosystem has grown over time, and it is becoming increasingly capable of statistical analysis.

It's the best balance of scale and sophistication (in terms od data processing).

Python places a premium on productivity and readability.

Python is used by programmers who want to do data analysis or use statistical techniques (and by devs that turn to data science)

Python scientific packages are available for data visualisation, machine learning, natural language processing, complex data analysis, and other purposes. All of these features combine to make Python an excellent tool for scientific computing and a viable alternative to commercial packages such as MatLab. The following are the most popular data science libraries and tools:

Pandas is a data manipulation and analysis library. Data structures and operations for manipulating numerical tables and time series are provided by the library.

NumPy: the foundational Python package for scientific computing, adding support for large, multi-dimensional arrays and matrices, as well as a large library of high-level mathematical functions to operate on these arrays.

SciPy is a Python library used by scientists, analysts, and engineers for scientific and technical computing.

Python has been widely adopted by the scientific community because it is a free, cross-platform, general-purpose, and high-level programming language. Python is valued by scientists for its precise and efficient syntax, relatively flat learning curve, and ability to integrate well with other languages (e.g., C/C++).

As a result of its popularity, Python scientific packages for data visualisation, machine learning, natural language processing, complex data analysis, and other purposes abound. All of these features combine to make Python an excellent tool for scientific computing and a viable

alternative to commercial packages such as MatLab.

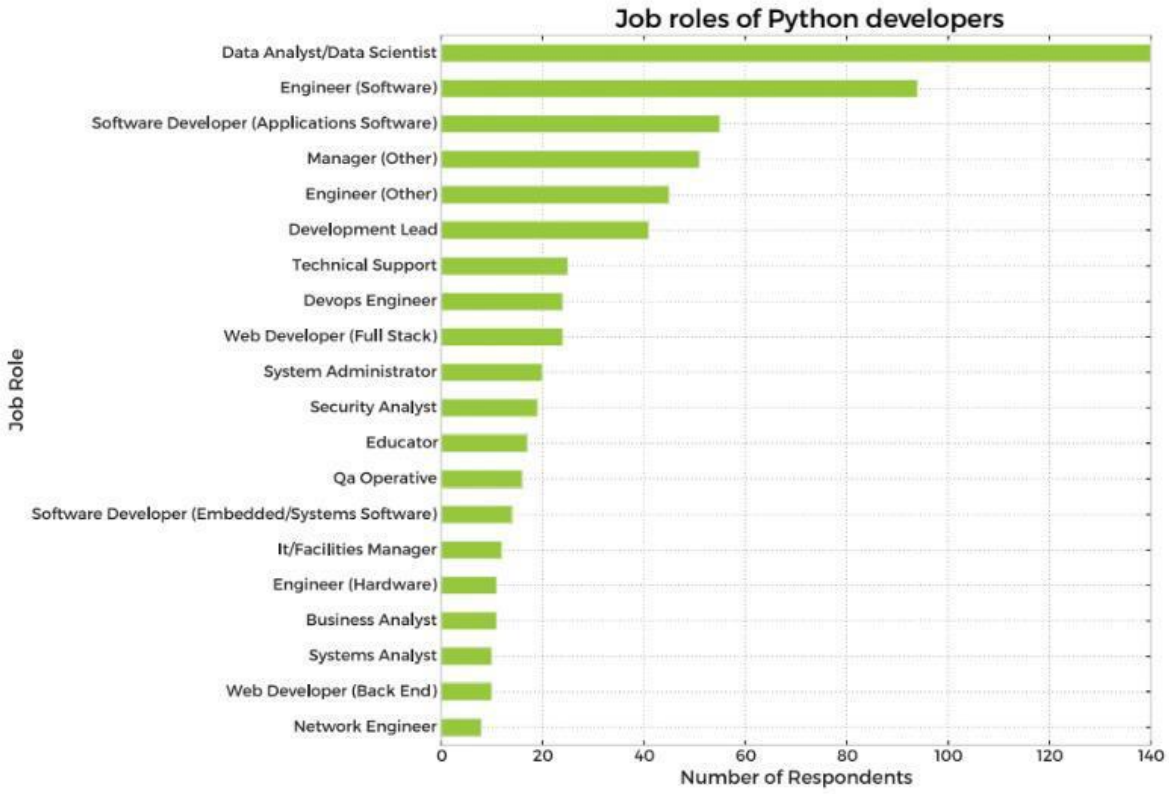


fig 1

Here's our list of the most popular Python scientific libraries and tools

Astropy

The Astropy Project is a collection of astronomy-related software packages. The core astropy package includes functionality geared toward professional astronomers and astrophysicists, but it may be useful to anyone developing astronomy software.

Biopython

Biopython is a set of free and open-source Python tools for computational biology and bioinformatics. It has classes for representing biological sequences and annotations, and it can read and write to a variety of file formats.

Cubes

Cubes is a lightweight Python framework and set of tools for developing reporting and analytical applications, OLAP, multidimensional analysis, and browsing aggregated data.

DEAP

DEAP is a framework for evolutionary computation that allows for rapid prototyping and testing of ideas. It includes the data structures and tools needed to implement the most common evolutionary computation techniques, including the genetic algorithm, genetic programming, evolution strategies, particle swarm optimization, differential evolution, and estimation of distribution algorithm.

SCOOP

SCOOP is a Python module that distributes concurrent parallel tasks on a variety of platforms, ranging from heterogeneous grids of workstations to supercomputers.

PsychoPy

PsychoPy is a Python package for creating experiments in neuroscience and experimental psychology. PsychoPy is intended to facilitate the presentation of stimuli and the collection of data for a variety of neuroscience, psychology, and psychophysics experiments.

Pandas

Pandas is a data manipulation and analysis library. Data structures and operations for manipulating numerical tables and time series are provided by the library.

Mlpy

Mlpy is a machine learning library built on top of the GNU Scientific Libraries NumPy/SciPy. Mlpy offers a wide range of machine learning methods for supervised and unsupervised problems, with the goal of striking a reasonable balance between modularity, maintainability, reproducibility, usability, and efficiency.

matplotlib

Matplotlib is a Python 2D plotting library that generates publication-quality figures in a variety of hardcopy and interactive formats across platforms. Matplotlib can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, and other graphics.

NumPy

NumPy is the foundational Python package for scientific computing, adding support for large, multi-dimensional arrays and matrices, as well as a large library of high-level mathematical functions for operating on these arrays.

NetworkX

NetworkX is a graph-analysis library that allows you to create, manipulate, and study the structure, dynamics, and functions of complex networks.

TomoPy

It's an open-source Python toolbox for processing tomographic data and reconstructing images. It's a collaborative framework for analysing synchrotron tomographic data, with the goal of bringing together the efforts of various facilities and beamlines performing similar tasks.

Theano

It's a Python library for numerical computation. It efficiently defines, optimises, and evaluates mathematical expressions involving multidimensional arrays.

SymPy

It's is a symbolic computation library with capabilities ranging from simple symbolic arithmetic to calculus, algebra, discrete mathematics, and quantum physics. It offers computer algebra capabilities as a standalone application, as a library to other applications, or as a live web application.

SciPy

It’s a Python library for scientific and technical computing that is used by scientists, analysts, and engineers. It modules support optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers, and other common tasks in science and engineering.

Scikit-learn

It’s a library for machine learning. It includes support vector machines, random forests, gradient boosting, k-means, and DBSCAN as classification, regression, and clustering algorithms, and is designed to work with the Python numerical and scientific libraries NumPy and SciPy.

Scikit-image

It's an image processing library. Segmentation, geometric transformations, colour space manipulation, analysis, filtering, morphology, feature recognition, and other activities are all covered by the algorithms.

ScientificPython

It's a collection of scientific computing modules in the form of a library. All of the following are supported: geometry, mathematical functions, statistics, physical units, IO, visualisation, and parallelization.

SageMath

It's mathematical software that incorporates, among other things, algebra, combinatorics, numerical mathematics, number theory, and calculus. It's written in Python and can handle procedural, functional, and object-oriented programming.

Veusz

It is a scientific plotting and graphing tool that produces high-quality graphs in popular vector formats like PDF, PostScript, and SVG.

Graph-tool

It is a module that allows you to manipulate and analyse graphs statistically.

SunPy

It is a Python-based data-analysis environment that specialises in providing the software required to analyse solar and heliospheric data.

Bokeh

It is a Python interactive visualisation library designed for presentation in modern web browsers. It is useful for anyone who wants to develop interactive graphs, dashboards, and data applications fast and easily. Its purpose is to give D3.js-style elegant and concise building of unique visuals while simultaneously delivering high-performance interactivity across very big or streaming datasets.

TensorFlow

It is an open source software library for machine learning that Google created to fulfil their demands for systems that can build and train neural networks to discover and analyse patterns and correlations in the same way that people learn and reason. It is now utilised in Google products for both research and production, frequently replacing DistBelief, its closed-source predecessor.

Nilearn

It is a Python module that makes statistical learning on Neuro.Imaging data quick and easy. It makes it simple to apply advanced machine learning, pattern recognition, and multivariate statistical approaches to neuroimaging data for applications including MVPA (Multi-Voxel Pattern Analysis), decoding, predictive modelling, functional connectivity, brain parcellations, and connectomes.

Dmelt

DataMelt, or DMelt, is a programme for numerical computation, statistics, large data volume analysis ("big data"), and scientific visualisation. The programme can be used in a variety of fields, including natural sciences, engineering, modelling, and financial market analysis. DMelt

can be used with a variety of scripting languages, including Python/Jython, BeanShell, Groovy, Ruby, and Java.

Python-weka-wrapper

Weka is a Java-based machine learning software suite created at New Zealand's University of Waikato. It offers a suite of data analysis and predictive modelling visualisation tools and algorithms, as well as graphical user interfaces for easy access to these capabilities. The python-weka-wrapper package allows you to use Python to run Weka algorithms and filters.

Dask

Dask is a parallel computing package for analytic computing that is made up of two parts: 1) Big Data collections such as parallel arrays, dataframes, and lists that extend common interfaces such as NumPy, Pandas, or Python iterators to larger-than-memory or distributed environments, and 2) dynamic task scheduling optimised for computation, optimised for interactive computational workloads.

Python Saves Time

Even the classic “Hello, world” program illustrates this point:

```
print("Hello, world")
```

For comparison, this is what the same program looks like in Java:

```
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello, world");
    }
}
```

All the Big Names Use Python

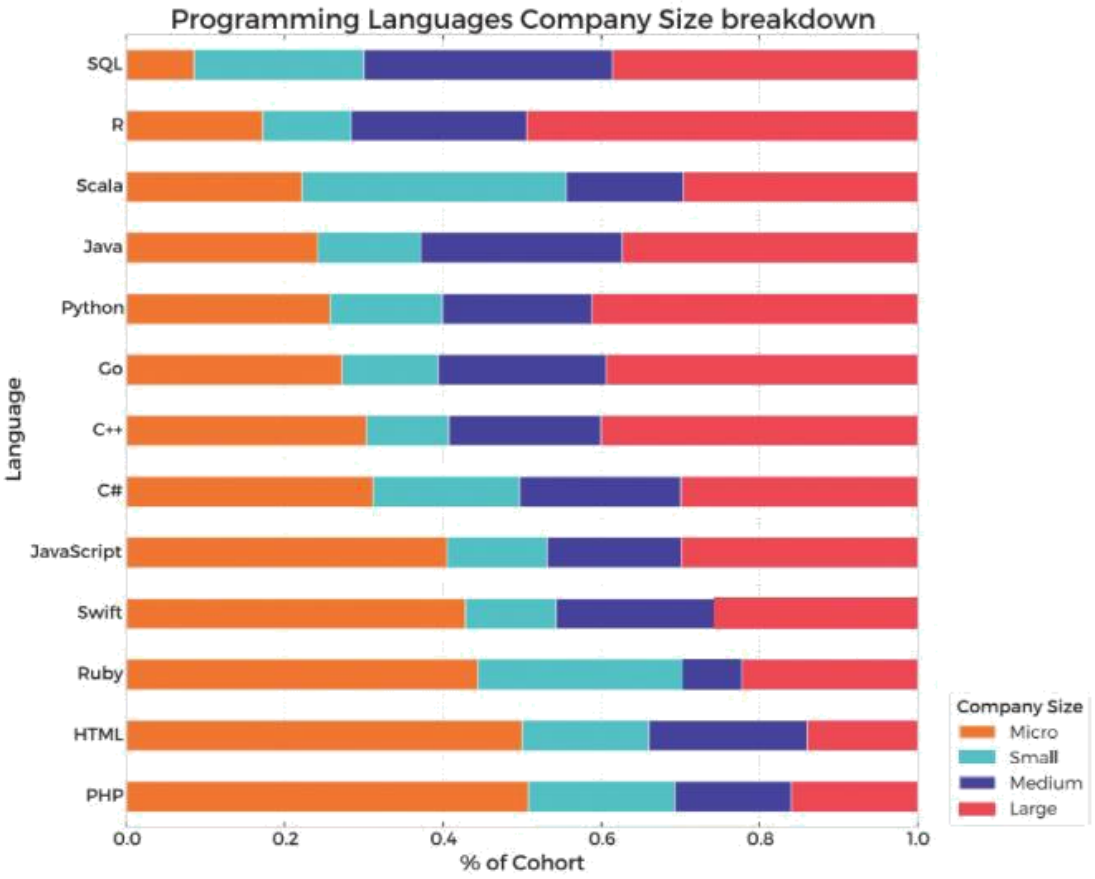


fig 2

Python Keywords and Identifier

Reserved words are referred to as keywords in Python.

A keyword cannot be used as a variable, function, or other identifier.

They are used to describe the syntax and organisation of the Python programming language.

Python uses case-sensitive keywords.

There are 33 keywords in Python 3.3. This number may fluctuate slightly over time.

Except for True, False, and None, all of the keywords are in lowercase and must be written exactly as they are. The complete list of keywords is provided below.

Keywords in Python programming language

False	class	finally	is	return
None	continue	for	lambda	try
True	def	from	nonlocal	while
and	del	global	not	with
as	elif	if	or	yield
assert	else	import	pass	
break	except	in	raise	

fig 3

In Python, an identifier is the name given to entities such as classes, functions, variables, and so on. It aids in distinguishing one entity from another.

Rules for writing identifiers

Identifiers can be a combination of letters in lowercase (a to z) or uppercase (A to Z) or digits (0 to 9) or an underscore (_). Names like myClass, var_1 and print_this_to_screen, all are valid example.

An identifier cannot start with a digit. 1variable is invalid, but variable1 is perfectly fine.

Keywords cannot be used as identifiers.

```
>>>global = 1
File"<interactive input>", line 1
global = 1
           ^
```

SyntaxError: invalid syntax

We cannot use special symbols like !, @, #, $, % etc. in our identifier.

```
>>> a@ =0
File"<interactive input>", line 1
    a@ =0
```

```
^
SyntaxError: invalid syntax
```

The length of the identifier is completely up to you.

Python is a general-purpose abnormal state programming dialect that is frequently used. Python features a dynamic sort architecture and programmed memory management, and it supports a variety of programming ideal models, such as object-oriented, fundamental, utilitarian, and procedural programming. It contains a sizable and well-rounded standard library. Some functional frameworks provide Python mediators, allowing Python code to operate on a variety.

Fig 4 Python Logo

4.System Analysis

SCOPE & METHODOLOGY

The goal of this SRS report is to separate the Intelligent Network Backup Tool's requirements and functionalities. The SRS will detail how our team and the client perceive the end product, as well as the features and functionality that it must possess. This record also contains a list of optional requirements that we want to fulfil but are not needed for the project's success.

This stage assesses the necessary requirements for Image Processing. For an organised manner of evaluating the prerequisites, a few processes are offered. Perceiving the concept of framework for a solid examination is the first step in dissecting the framework's requirements, and all situations are described to better comprehend the dataset's inquiry.

INTENDED AUDIENCE AND READING SUGGESTIONS

Extension engineers, directors, clients, analysts, and documentation journalists should use this record. This paper looks into plan and execution needs, as well as conditions, framework highlights, external interface requirements, and other non-functional requirements.

IDENTIFICATION OF NEEDS

Understanding how a commercial enterprise or group is performing in the market is the first and most crucial necessity, as well as how to defeat their competitors in the market.

In order to do so, we must investigate our data in light of all available variables.

FEASIBILITY STUDY

A credibility examiner expected to fairly and objectively identify the strengths and weaknesses of an existing business or projected course, as well as the opportunities and challenges that nature presents, the advantages that must be realised, and the long-term chances for growth. In the simplest terms, the incurred major injury necessary and the motivation to the fulfilment are the two factors for determining believability.

An inside and out planned feasibility study should include a written history of the business or venture, a description of the thing or organisation, accounting explanations, operational and organisational objectives of enthusiasm, publicising examination and game plans, budgetary data, genuine necessities, and cost obligations. Overall, probability looks at going before specific change and wander utilisation. Achievability can be divided into three categories: economic feasibility, technical feasibility, and operational feasibility.

Economical feasibility

Apart from a large collection of other organization's reports, the electronic structure should totally control the current existing system's data stream and technique and should create all of the manual structure's reports. It should be finished as an electronic application with its own web server and database server. A segment of the associated trades advances in a variety of ranges. Open source programming such as TOMCAT, JAVA, MySQL, and Linux are used to keep costs down for the consumer. There is no need for an extraordinary wander to manage the instrument.

Technical feasibility

The most difficult aspect of a believability analysis is surveying the specific probability. This is because, as of now, there isn't a lot of detail in the system's layout, making it difficult to get to concerns like execution, costs (according to the type of development to be passed on), and so on. When conducting a specialised examination, a number of factors must be taken into account. Recognize the many steps that must be followed in the suggested system. We should be clear about the developments that will be required for the transition to the new system before we begin our journey. Check to check if the company has all of the required innovations.

Is it possible to get the necessary training through the affiliation?

Is the utmost sufficient in this case?

For example, "Will the current printer be able to manage the new reports and structures required by the new system?"

Operational feasibility

Proposed excursions will only be profitable if they can be converted into information systems that fulfil the organization's operational needs. Simply expressed, this probability test determines whether or not the structure will work after it is constructed and presented. Is there a genuine barrier to implementation? Here are some questions to help you assess a wander's operational viability.

• Is there enough assistance from customers for the wander from the organisation? • There may be opposition if the current structure is particularly cherished and used to the point where people are unable to see the reasons for change.

• Are the current business approaches customer-friendly? Users may embrace a change that will result in more operational and supporting systems if they aren't already.

• Has the customer agreed to the orchestrating and changing of the wander? Early commitment reduces the likelihood of structure impenetrability.

Economical feasibility

Apart from a large collection of other organization's reports, the electronic structure should totally control the current existing system's data stream and technique and should create all of the manual structure's reports. It should be finished as an electronic application with its own web server and database server. A segment of the associated trades advances in a variety of ranges. Open source programming such as TOMCAT, JAVA, MySQL, and Linux are used to keep costs down for the consumer. There is no need for an extraordinary wander to manage the instrument.

Technical feasibility

The most difficult aspect of a believability analysis is surveying the specific probability. This is because, as of now, there isn't a lot of detail in the system's layout, making it difficult to get to concerns like execution, costs (according to the type of development to be passed on), and so on. When conducting a specialised examination, a number of factors must be taken into account. Recognize the many steps that must be followed in the suggested system. We should be clear about the developments that will be required for the transition to the new system before we begin

our journey. Check to check if the company has all of the required innovations. Is it possible to get the necessary training through the affiliation?

Is the utmost sufficient in this case?

For example, "Will the current printer be able to manage the new reports and structures required by the new system?"

Operational feasibility

Proposed excursions will only be profitable if they can be converted into information systems that fulfil the organization's operational needs. Simply expressed, this probability test determines whether or not the structure will work after it is constructed and presented. Is there a real roadblock to implementation? Here are some questions to consider when evaluating the operational viability of a wander.

• Is there enough assistance from customers for the wander from the organisation? There may be opposition if the current structure is particularly cherished and used to the point where people are unable to see the reasons for change.

• Are the current business methodologies customer-qualified? If they are not, users may appreciate a change that will result in more operational and supporting systems.

• Is the customer involved in the orchestrating and changing of the wander? Early commitment reduces the likelihood of structure impenetrability.

SOFTWARE REQUIREMENTS

Operating System	:	Windows
Framework	:	Jupyter
Language	:	Python
IDE	:	Anaconda

HARDWARE REQUIREMENTS

Processor	:	Pentium 4
Hard disc	:	500GB
RAM	:	4GB

System includes all standard accessories such as a monitor, keyboard, mouse, and so on.

5. DESIGN:

The System Design Document details the system requirements, operating environment, system and subsystem architecture, files and database design, input formats, output layouts, human-machine interfaces, detailed design, processing logic, and external interfaces.

This section gives a non-technical narrative description of the system. A high-level system architecture diagram with a subsystem breakdown, if applicable, should be included. External system interfaces should be depicted in high-level system architecture or subsystem diagrams, if applicable. Provide a high-level context diagram for the system and subsystems, if applicable. Consult the RTM in the FRD to determine the functional requirements' allocation in this design document.

This section comprises any assumptions made by the project team in establishing the system design, as well as any constraints in the system design (such as resource use versus productivity or conflicts with other systems).

The important points of contact for the information system development effort's organisation code and title (and alternates if necessary). The Project Manager, System Proponent, User Organization, Quality Assurance (QA) Manager, Security Manager, and Configuration Manager should all be included in this list.

SYSTEM ARCHITECTURE

UML Diagrams

Selecting, visualising, designing, and describing programming structures is done using UML. The UML is a visual notation for creating programming plans. It can also be used to show non-programming structures, such as a process stream in a gathering unit.

UML is a set of tools that may be used to produce code in a variety of languages using UML graphs, rather than a programming language. UML provides a link to question-based testing and outlines. UML is expected to play a key role in representing a structure's trade viewpoints.

Use case Diagram:

The use case graph is used to show the directness of the structure.

This diagram depicts the flow of use cases, performing pros, and their relationships.

This diagram could be utilized to deal with the structure's static viewpoint.

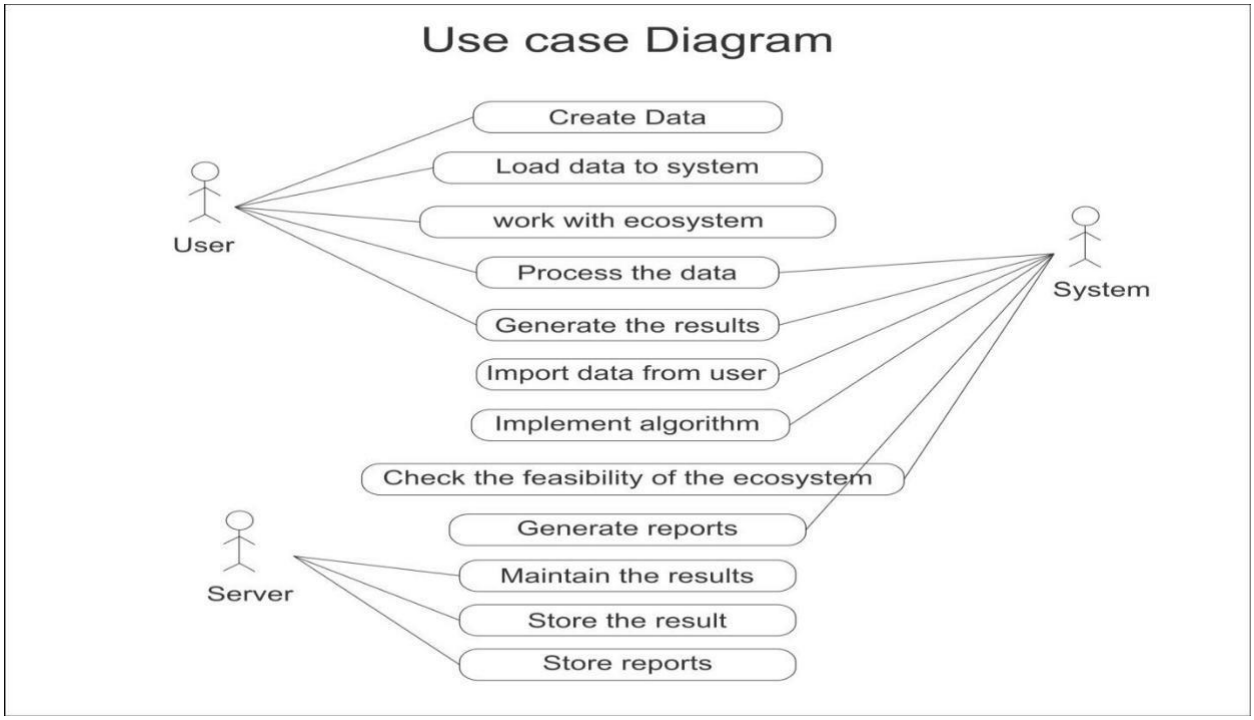


fig:5 :use case diagram

Customer, structure, client, server, Hadoop, and data cleansing are the performing professionals in the diagram above. The data is sent from the client to the system, which separates it into squares before sending it to Hadoop. Then Hadoop performs data cleaning, which is simply connecting and repairing data, and the results are secure. These results can be viewed with Hadoop and saved in a server for future use. The customer can generate reports based on the obtained results.

Class Diagram:

The most often used UML layout component is the class graph.

It is concerned with the structure's static course of operation.

It solidifies the strategy of classes, interfaces, collaborative efforts, and affiliations.

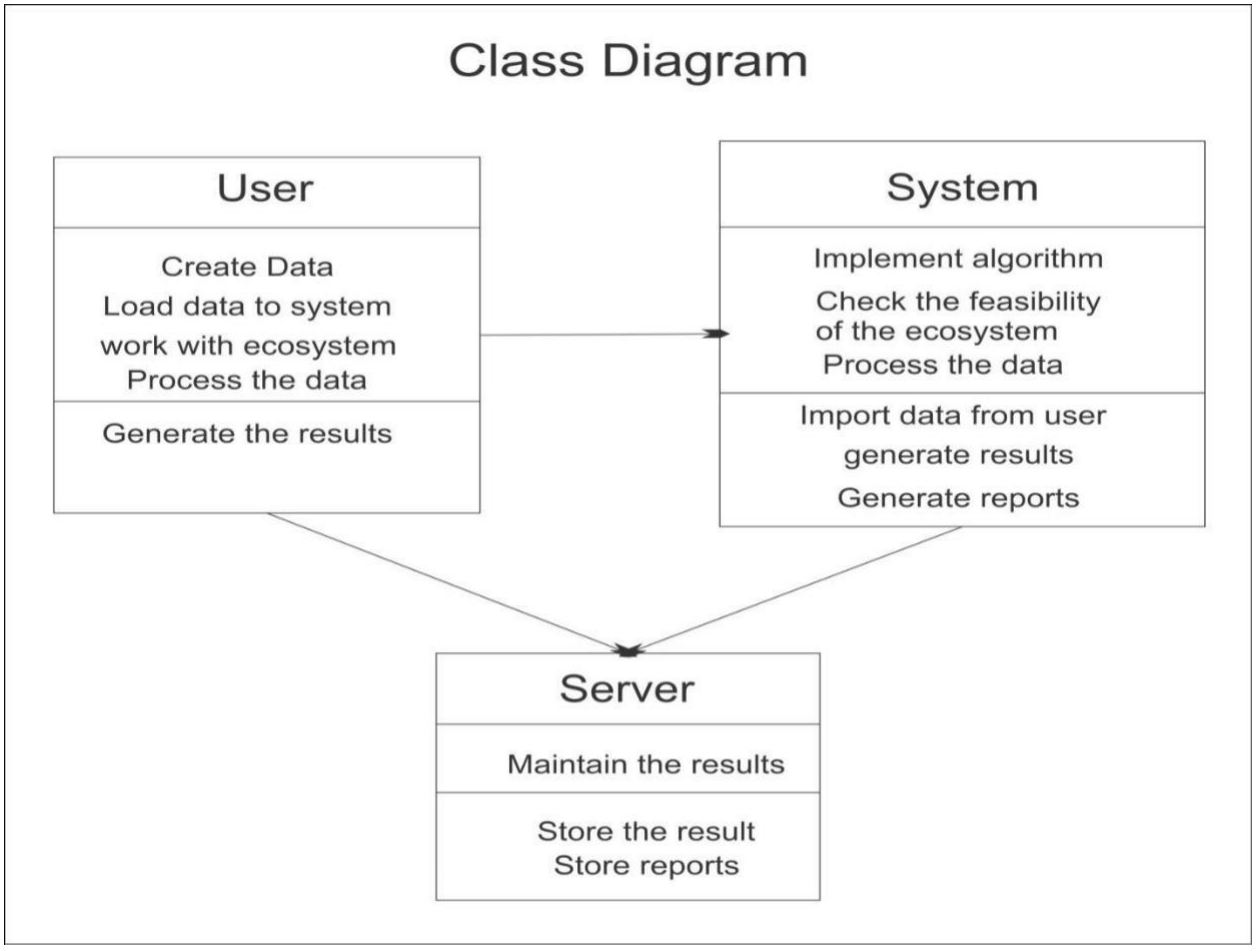


fig:6 :class diagram

In the above class diagram, the relationship between each of the classes is depicted as a dependency. Furthermore, the operations that are carried out in each class are displayed in a similar manner.

Sequence Diagram:

This is a cooperative design that focuses on message requests at specific times.

It consists of a set of parts as well as the messages sent and received by the instance of parts.

This diagram depicts the structure's dynamic perspective.

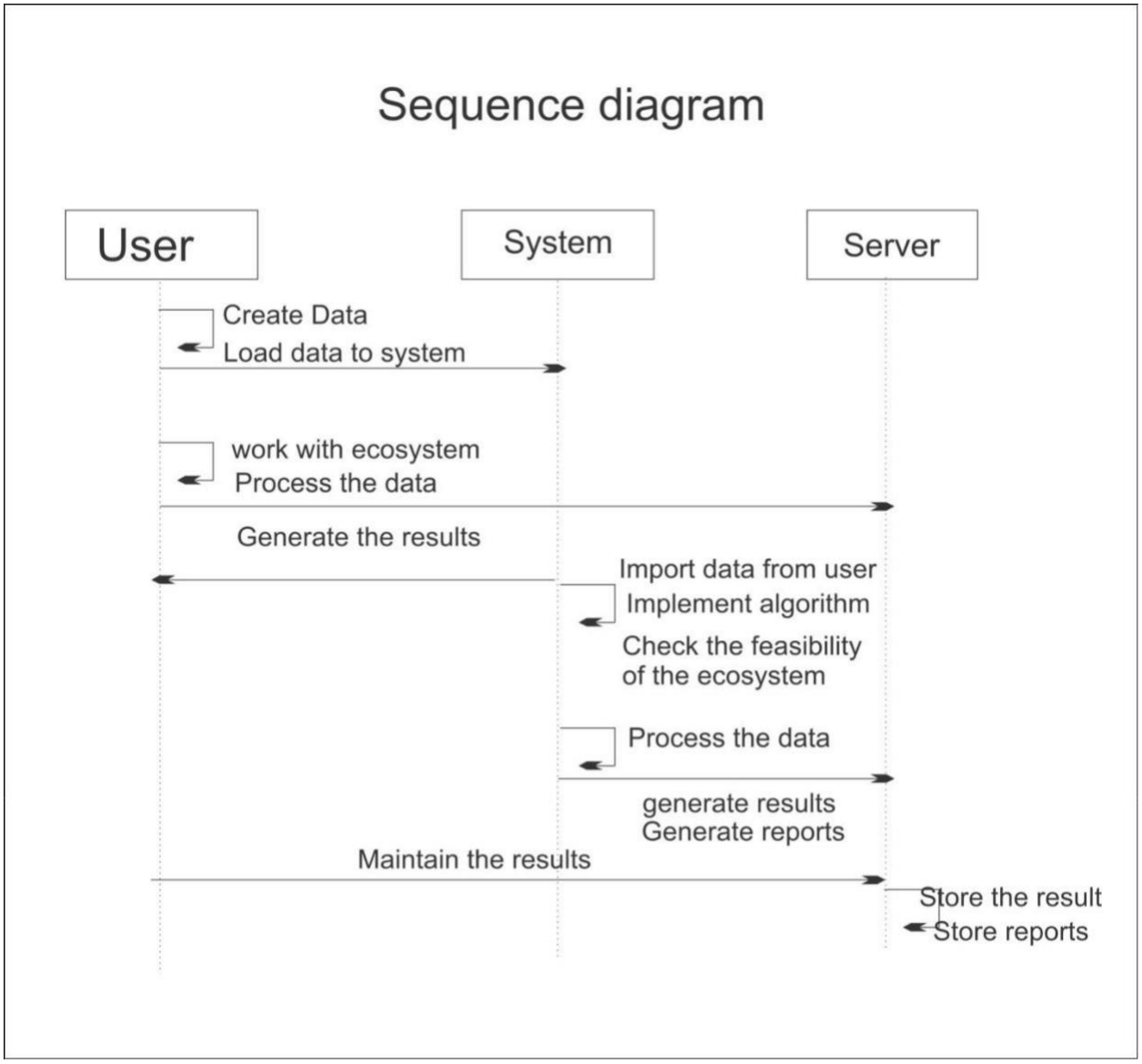


fig:7 :sequence diagram

The term "success outline" refers to the order in which questions are communicated. Five articles in the graph above are collaborating with one another. A vertical dashed line runs through each protest, indicating the presence of a query over an indeterminate time span. This graph also includes a tall, thin rectangle termed the centre of control, which depicts the timeframe in which a protest is carrying out an activity, either directly or indirectly.

Collaboration Diagram:

This is a help format that concentrates on the main relationship between articles that send and receive messages. It is made up of a number of pieces, connectors that connect them, and messages delivered and received by the parts. This diagram is designed to illustrate the framework's dynamic viewpoint.

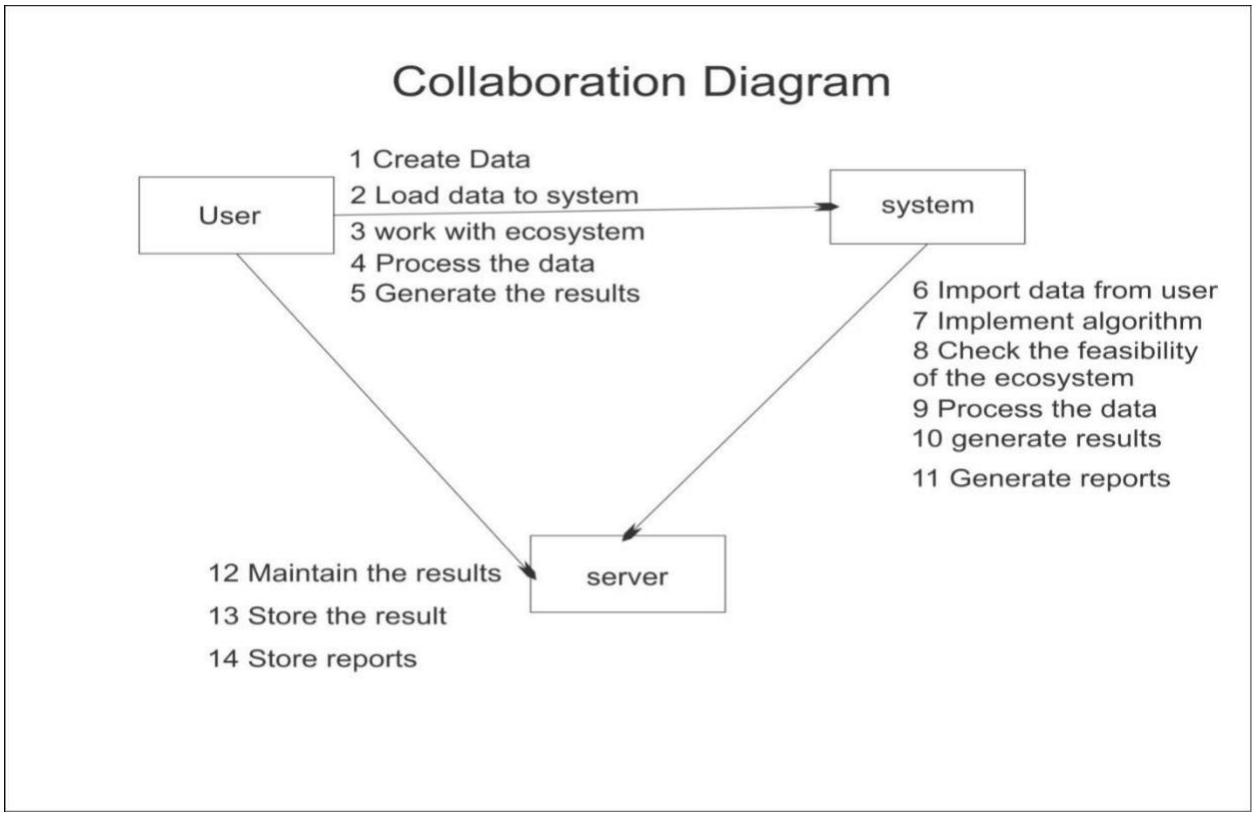


fig:8: collabration diagram

The joint effort outline includes articles, a method, and an arrangement number. There are five questions in the graph above: customer, client, framework, Hadoop, and server. The components are linked together by means of a path. A succession number represents a message's time request.

State Chart Diagram:

The game plan of states, occasions, and exercises is represented by the state graph.

This graph is notable for focusing on the interface's lead, class, and made effort.

The primary focus of state outline is to demonstrate the occasion will sort out the request's lead.

The dynamic perspective of the framework is depicted by the state follows diagram.

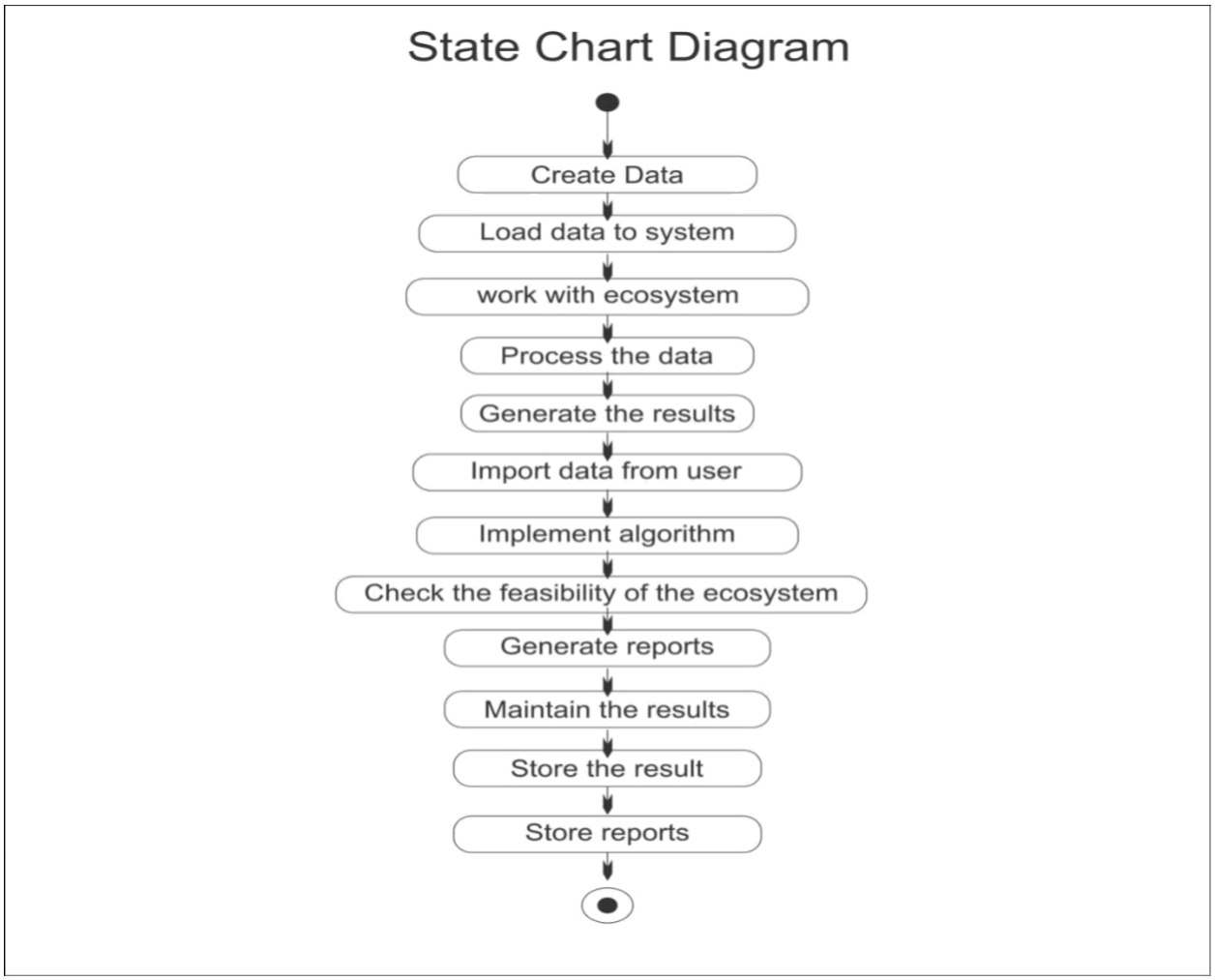


fig:9 :state chart diagram

A state outline graph is made up of two parts: situation of affairs and advancement States refer to conditions that occur during the course of a person's life. In Smart Draw, we can simply sketch a state by using a rectangle with modified corners. Change is a powerful bolt that shatters the status quo. path that a question takes between various conditions. Name the change, as well as the event that caused it and the action that followed.

Component Diagram:

Segment is a required component of part format. This diagram demonstrates the piece's internal parts, connectors, and ports. Duplicates of internal parts are created when a section is instantiated.

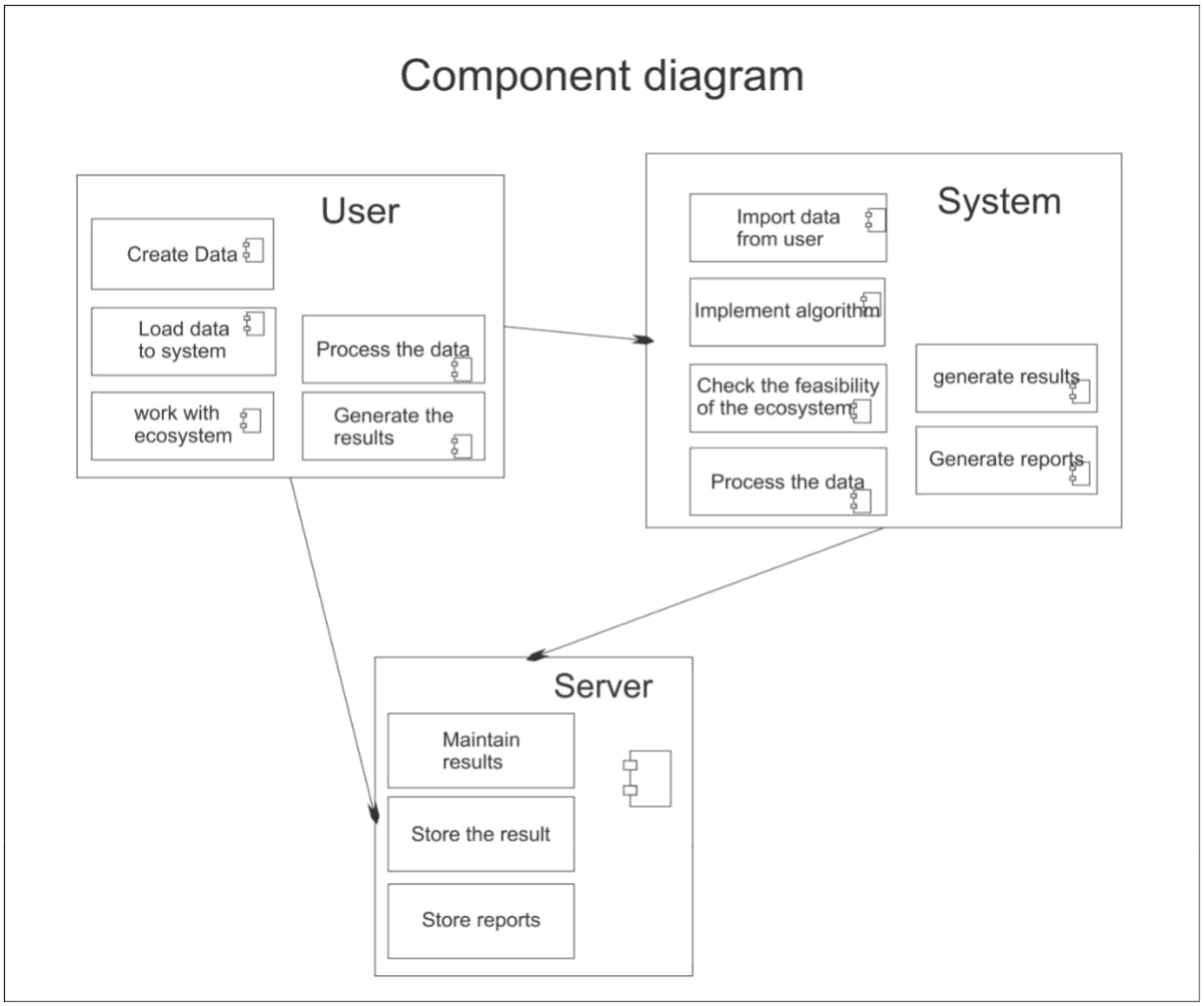


fig:10 :component diagram

A segment is used to speak to a part outline. A part is a component of the framework that is physically present. It is referred to as a rectangle with a tab. The internal management of the venture is depicted by the part outline. The data is then sent to Hadoop, where sqoop is used to clean the data and hive is used to generate reports.

Deployment Diagram:

A middle point is the most important component in game-plan layout. Sending storyline tends to the approach of emphasising on focuses and their relationships with one another. The area diagram, which is one focus purpose evidently of activity format that typically includes at least one part, identifies the sending outline. This outline is also essential for tending to the framework's static perspective.

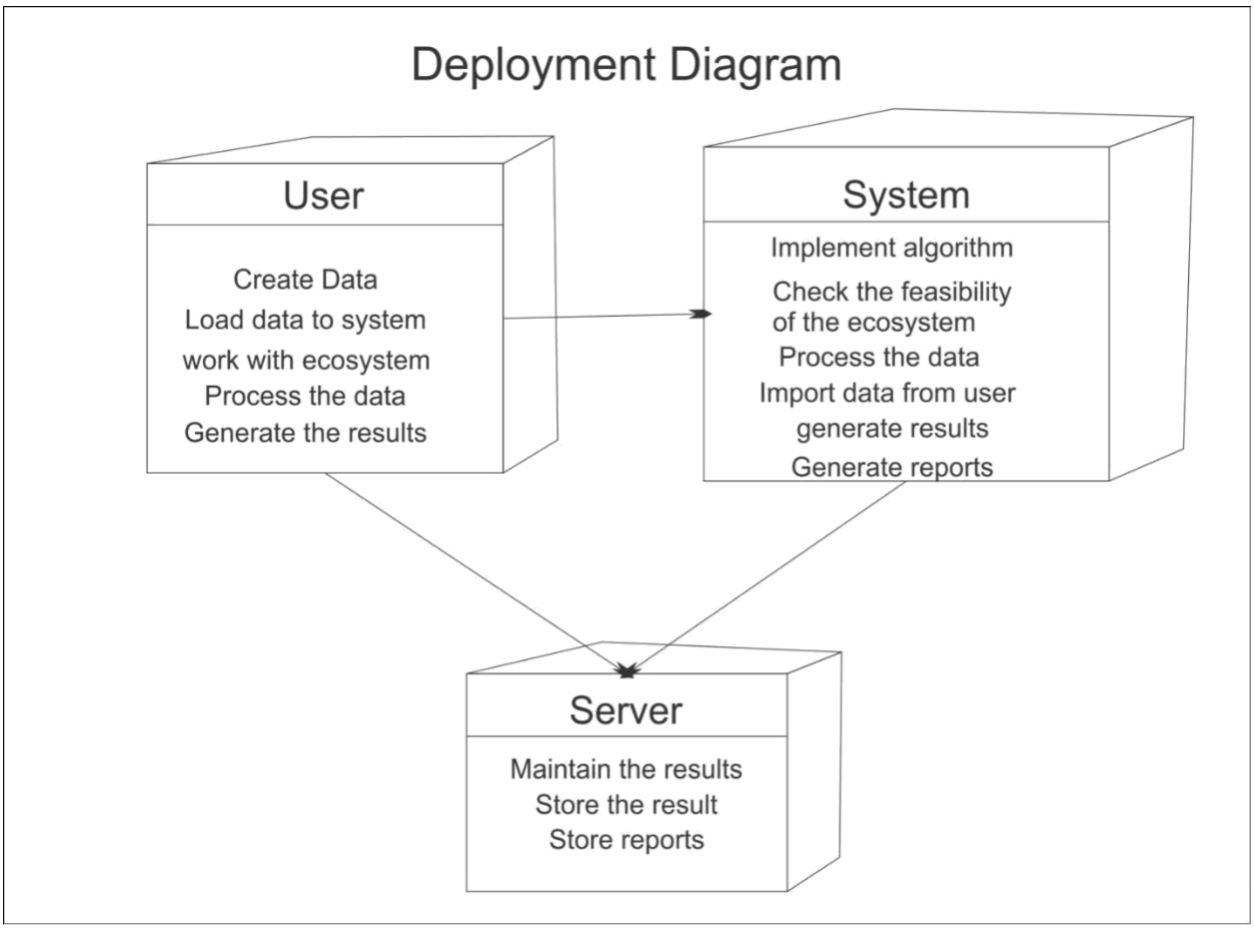


fig:11 :deployment diagram

Hub is used to communicate with an arrangement graph. A physical item that runs code is referred to as a hub. They are also used to depict hub runtime management. The data is then sent to Hadoop, where sqoop is used to clean the data and hive is used to generate reports.

5.4 DATA FLOW DIAGRAMS

An information stream design (DFD) is a graphical depiction of the "stream" of information in a data framework, displaying its strategy edges. A DFD is typically used as a preliminary walk to have a better understanding of the framework, which may later be clarified. DFDs can also be used to represent data preparation. The sort of data that will be sensed and yielded from the structure, as well as where the information will begin and end, and how it will be safeguarded, are all specified in a DFD. It doesn't show information on process organisation or if the approach

will be implemented in a sequential or parallel manner.

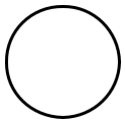
DFD Symbols:

There are four symbols in the DFD.

A source or destination of system data is defined by a square.

Data flow is shown by an arrow. It's the conduit through which information travels.

The process of transforming incoming data flow into exiting data flow is represented by a circle.



A data storage, data at rest, or a transitory data repository is an open rectangle.

Level 0: System input/ output level The system's broad boundaries, comprising input and output flows, as well as important processes, are described in a level 0 DFD.

fig 12 level 0 dfd

Context Diagram is another name for DFD Level 0. It is a critical examination of the entire structure or process that has failed or appeared. It must be a first watch, exhibiting the framework as a unique state handle with its connections to exterior substances.

Level 1: Sub system level data flow

The corresponding level of enthusiasm for the data stream between subsystems is defined by Level 1 DFD. The Level 1 DFD depicts the system's division into sub-structures (shapes), each of which supervises at least one of the data streams to or from an outside entity, and which together offer the majority of the system's utility.

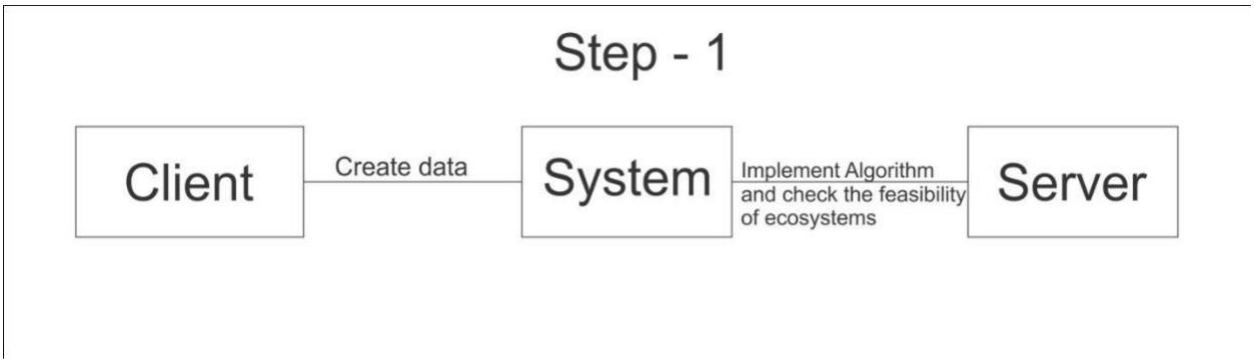


fig 13 level 1 dfd

Level 2: File level detail data flow

From various perspectives, the examination of possibility and danger is linked here.

The level 2 DFD explains the system's operation at the most fundamental level.

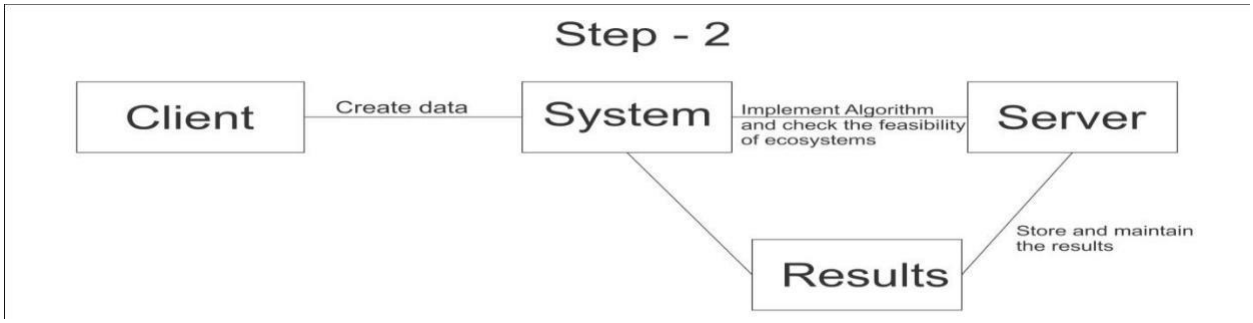


fig 14 level 2 dfd

6. IMPLEMENTATION

Modules

Dataset:

Diabetic patient distribution- We created a model to predict diabetes, but the dataset was slightly imbalanced, with around 500 classes labeled as 0 means no diabetes and 268 labeled as 1 means diabetic.

Data Preprocessing- The most important process is data preprocessing. Most healthcare-related data contains missing values and other impurities that can reduce data effectiveness. Data preprocessing is performed to improve the quality and effectiveness of the results obtained from the mining process. To effectively apply Machine Learning Techniques to a dataset, this process is required for accurate results and successful prediction. We need to preprocess the Pima Indian diabetes dataset in two steps.

Missing Values Removal- All instances having a value of zero should be removed (0). There is no such thing as a value of zero. As a result, this instance isn't valid any longer. By removing irrelevant features/instances, we create a feature subset, and this process is known as feature subset selection. This reduces the dimension of the data and allows us to work faster.

Data splitting- After cleaning the data, the data is normalised for training and testing the model. When the data is spitted, we train the algorithm on the training data set while keeping the test data set aside. This training process will generate a training model based on logic and algorithms, as well as the values of the features in the training data. The goal of normalisation is to bring all attributes onto the same scale.

Apply Machine Learning- Once the data is ready, we use the Machine Learning Technique. To predict diabetes, we employ various classification and ensemble techniques. The methods were tested on a diabetes dataset from Pima Indians. The main goal is to use Machine Learning Techniques to analyse the performance of these methods and determine their accuracy, as well as to identify the responsible/important feature that plays a significant role in prediction.

MODEL BUILDING

This is the most important phase, which includes the development of a model for diabetes prediction. In this, we implemented various machine learning algorithms for diabetes prediction, as discussed above. Procedure of Proposed Methodology-

Step 1: Import the necessary libraries and the diabetes dataset.

Step 2: Pre-process the data to remove any missing information.

Step 3: Perform a percentage split of 80 percent to divide the dataset into a Training set and a Test set.

Step 4: Choose a machine learning algorithm, such as K-Nearest Neighbor, Support Vector Machine, Decision Tree, Logistic Regression, Random Forest, or Gradient Boosting.

Step 5: Using the training set, create a classifier model for the mentioned machine learning algorithm.

Step 6: Using the test set, test the Classifier model for the aforementioned machine learning algorithm.

Step 7: Conduct a comparison Evaluation of the results of each classifier's experimental performance.

Step 8: After analysing various metrics, select the best performing algorithm.

Code:

Jupyter Untitled2 Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run Code

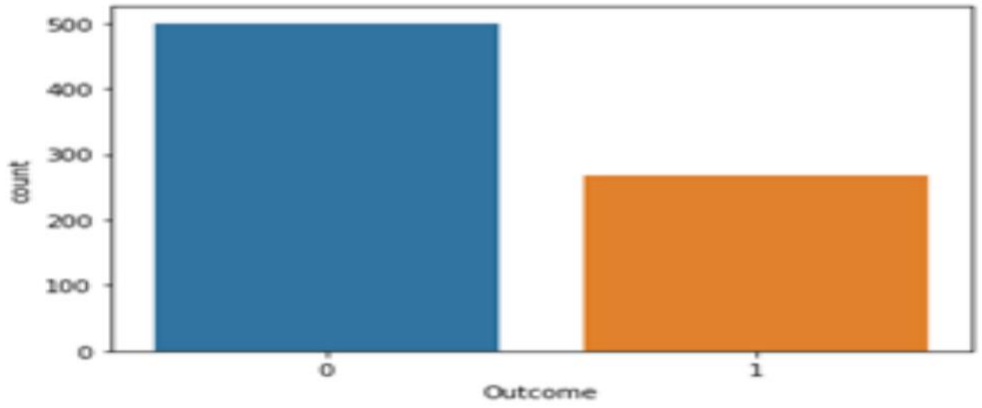
In [17]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [18]:
```
dataset = pd.read_csv('C:/Users/user/diabetes.csv')

dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
Pregnancies                 768 non-null int64
Glucose                     768 non-null int64
BloodPressure               768 non-null int64
SkinThickness               768 non-null int64
Insulin                     768 non-null int64
BMI                         768 non-null float64
DiabetesPedigreeFunction    768 non-null float64
Age                         768 non-null int64
Outcome                     768 non-null int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

```
sns.countplot(x = 'Outcome',data = dataset)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x14e3bf10860>
```



Jupyter Untitled2 Last Checkpoint: an hour ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help

Run Code

In [19]:
```
dataset.head()
```

Out[19]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

In [20]:
```
dataset.describe().T
```

Out[20]:

	count	mean	std	min	25%	50%	75%	max
Pregnancies	768.0	3.845052	3.369578	0.000	1.00000	3.0000	6.00000	17.00
Glucose	768.0	120.894531	31.972618	0.000	99.00000	117.0000	140.25000	199.00
BloodPressure	768.0	69.105469	19.355807	0.000	62.00000	72.0000	80.00000	122.00
SkinThickness	768.0	20.536458	15.952218	0.000	0.00000	23.0000	32.00000	99.00
Insulin	768.0	79.799479	115.244002	0.000	0.00000	30.5000	127.25000	846.00
BMI	768.0	31.992578	7.884160	0.000	27.30000	32.0000	36.60000	67.10
DiabetesPedigreeFunction	768.0	0.471876	0.331329	0.078	0.24375	0.3725	0.62625	2.42
Age	768.0	33.240885	11.760232	21.000	24.00000	29.0000	41.00000	81.00
Outcome	768.0	0.348958	0.476951	0.000	0.00000	0.0000	1.00000	1.00

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [21]:
```
dataset_new = dataset

dataset_new[["Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI"]] = dataset_new[["Glucose", "BloodPressure", "SkinThi
```

In [13]:
```
dataset_new.isnull().sum()
```

Out[13]:
```
Pregnancies                 0
Glucose                     5
BloodPressure              35
SkinThickness             227
Insulin                   374
BMI                        11
DiabetesPedigreeFunction    0
Age                         0
Diabetes                    0
dtype: int64
```

In [22]:
```
dataset_new["Glucose"].fillna(dataset_new["Glucose"].mean(), inplace = True)
dataset_new["BloodPressure"].fillna(dataset_new["BloodPressure"].mean(), inplace = True)
dataset_new["SkinThickness"].fillna(dataset_new["SkinThickness"].mean(), inplace = True)
dataset_new["Insulin"].fillna(dataset_new["Insulin"].mean(), inplace = True)
dataset_new["BMI"].fillna(dataset_new["BMI"].mean(), inplace = True)
```


In [10]:
```
from sklearn.preprocessing import MinMaxScaler
sc = MinMaxScaler(feature_range = (0, 1))
dataset_scaled = sc.fit_transform(dataset_new)
```

In [11]:
```
dataset_scaled = pd.DataFrame(dataset_scaled)
X = dataset_scaled.iloc[:, [1, 4, 5, 7]].values
Y = dataset_scaled.iloc[:, 8].values
```

In [12]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.20, random_state = 42, stratify = dataset_new['Outcome']
```

In [13]:
```
from sklearn.ensemble import RandomForestClassifier
ranfor = RandomForestClassifier(n_estimators = 11, criterion = 'entropy', random_state = 42)
ranfor.fit(X_train, Y_train)
Y_pred_ranfor = ranfor.predict(X_test)
```

In [14]:
```
from sklearn.svm import SVC
svc = SVC(kernel = 'linear', random_state = 42)
svc.fit(X_train, Y_train)
Y_pred_svc = svc.predict(X_test)
```

In [16]:
```
from sklearn.metrics import accuracy_score
accuracy_ranfor = accuracy_score(Y_test, Y_pred_ranfor)
accuracy_svc = accuracy_score(Y_test, Y_pred_svc)
```

In [17]:
```
print("Support Vector Classifier: " + str(accuracy_svc * 100))
print("Random Forest: " + str(accuracy_ranfor * 100))
```

```
Support Vector Classifier: 73.37662337662337
Random Forest: 75.97402597402598
```

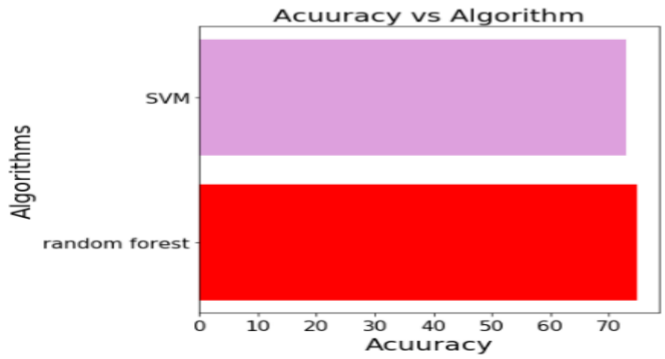
In [18]:
```
from sklearn.metrics import classification_report
print(classification_report(Y_test, Y_pred_svc))
```

```
              precision    recall  f1-score   support

         0.0       0.77      0.85      0.81       100
         1.0       0.65      0.52      0.58        54

    accuracy                           0.73       154
   macro avg       0.71      0.68      0.69       154
weighted avg       0.73      0.73      0.73       154
```

In [19]:
```
data = {'Algorithms':['Random Forest Classifier', 'SVM'],'Accuracy':[accuracy_ranfor,accuracy_svc]}
df = pd.DataFrame(data)
print(df)
```

```
                    Algorithms  Accuracy
0  Random Forest Classifier  0.759740
1                        SVM  0.733766
```

In [58]:
```
abc=[75,73]
plt.figure(figsize=(6,6))
colors=['red','plum']
labels=['random forest','SVM']
plt.barh(labels,abc,color=colors)
plt.xlabel("Acuuracy",fontsize=20)
plt.ylabel("Algorithms",fontsize=20)
plt.title("Acuuracy vs Algorithm",fontsize=20)
plt.xticks(rotation=1,fontsize=15)
plt.yticks(fontsize=15)
plt.show()
```



7. Testing

7.1 INTRODUCTION TO TESTING

A process for discovering faults in a programme is called testing. Programming testing is a crucial part of programming quality assurance, and it entails a thorough examination of the determination, outline, and coding. The cost of a product failure, as well as the expanding adoption of programming as a framework component, are driving variables that we arranged through testing. The process of executing a programme with the goal of finding a problem is known as testing. The test strategy for programming and other constructed items can be as challenging as the item's underlying outline. It is the most important quality metric used in programming development. During testing, the programme is run through a series of experiments, and the program's output for each experiment is evaluated to determine whether the programme is performing as expected.

TESTING STRATEGIES

Programming testing combines the design of programming experiments into a well-thought-out series of actions that leads to successful product improvement. The procedure serves as a roadmap for completing the tasks, when they should be completed, and how much work, time, and resources will be necessary. Test planning, experiment design, test execution, and test outcome collecting and evaluation are all part of the procedure. The procedure serves as both a guide for the specialist and a checklist for the chief. Advances must be quantifiable due to time limits, and issues must be identified as soon as possible.

Keeping in mind the end goal of ensuring that the framework is free of flaws, the various levels of testing techniques that are connected at various stages of programming development are:

Unit Testing

Unit testing is performed as individual modules are completed and made executable. It is limited to the requirements of the planner. It concentrates testing on the programming module or capability. It concentrates on the interior rationale and information structures. When a module is built with high union, it is rearranged.

- The number of experiments is reduced.
- Enables errors to be foreseen and revealed in a more effective manner.

Black Box Testing

It is also known as functional testing. A product testing strategy in which the analyzer is unaware of the inner workings of the thing being tested. For example, in a discovery test on a product outline, the analyzer only knows the information sources and what the expected results should be, but not how the programme arrives at those results. The analyzer never inspects the programming code and requires no additional knowledge of the programme other than its determinations. In this system, some experiments are constructed as information conditions that fully execute all of the program's practical requirements. The following categories of mistakes were discovered as a result of this testing:

- Capabilities that are incorrect or absent
- Mistakes in the user interface
- Errors in the information structure or unauthorised access to the database
- Errors in performance
- Errors in the beginning and at the end

Only the yield is checked for correctness in this test.

White Box testing

Glass box, structural, clear box, and open box testing are all terms used to describe this type of testing. A product testing technique in which test information is selected using express knowledge of the inner workings of the thing being tested. White box testing, as opposed to discovery testing, uses specific learning of programming code to inspect outputs. The test is precise only if the analyzer understands what the programme is supposed to do. He or she would then be able to see if the programme deviates from its intended goal. White box testing does not represent mistakes caused by oversight, and all obvious code should be discernible as well. It is necessary to do both white box and discovery tests for a comprehensive programming exam.

Experiments on each module's reasoning are created by drawing stream diagrams for that module, and reasonable alternatives are examined in each case. In the following examples, it was utilised to create the experiments:

In the following cases, it was used to create experiments:

• Ensure that each and every free path has been taken.

• Execute each and every wise decision on both truthful and false sides.

Integration Testing

Test in coordination ensures that the product and its subsystems work in concert. It tests the interface of a large number of modules to ensure that they work properly when synchronised. It is defined as a methodical approach to product engineering development. While the reconciliation is taking place, Conduct lead tests to identify any interface-related issues. Its purpose is to take unit-tested modules and create a programme structure based on the recommendations.

Non-incremental integration testing and incremental integration testing are two approaches to a test of integration.

System Testing

Internal testing of the entire framework prior to delivery to the client is referred to as framework testing. Its purpose is to satisfy the client by ensuring that the framework satisfies all of their criteria. With the use of a specific report, this testing evaluates the framework's functionality from the client's perspective. It does not necessitate any in-depth knowledge of the framework, such as the plan or structure of the code.

It has both utilitarian and non-utilitarian utilization/item zones. All of the major types of testing are encased in Framework Testing, which is referred to as a super arrangement of a wide variety of testing. Despite the fact that the focus on various types of testing may fluctuate depending on the item, association techniques, sequence of events, and requirements. Framework

testing is the first step in true testing, in which you test the entire object rather than just a module or highlight.

Acceptance Testing

Acknowledgement testing is a type of testing used to determine if the product framework meets needs. The primary purpose of this test is to see if the framework conforms with business needs and meets the requirements for end-user delivery. It is a type of pre-conveyance testing in which the entire framework is tested on real data at the customer's site to find flaws. The outcomes of the acknowledgment test bodies of evidence are compared to the standard ones after they are conducted against the test data or employing an acknowledgment test content.

The activities for the acknowledgement test are completed in stages. The necessary tests are carried out first, and if the test results are acceptable, more complex cases are carried out afterwards.

7.2 TEST APPROACH

A test methodology describes how a company's test system will be used and how testing will be carried out. One of the most essential variables in the effectiveness of the test effort and the accuracy of the test designs and gauges is the choice of test methodology or test techniques.

Testing should be possible from both the top down and the bottom up.

Bottom up Approach

Testing might begin with the smallest and most basic level modules and progress through them one by one. This approach directs testing from sub module to principal module, and if the primary module is not formed, it is recreated using a transitory programme called DRIVERS. When the lower-level modules are evaluated, the focus switches to the higher-level modules that utilise the lower-level ones. These are examined independently before being attached to the lower-level modules that have already been reviewed.

Top down Approach

In this technique, testing is directed from the core module to the sub module. If the sub module isn't built, a temporary programme called STUB is used to imitate it. Upper-level modules are

the starting point for this type of testing. Stubs are created because the nitty gritty exercises are frequently performed in the lower level schedules. A stub is a lower-level module that calls a higher-level module &, when successfully completed, returns a message to the caller module indicating that the required association was made.

7.3 VALIDATION

The method for evaluating programming during the development process or at the end of the development phase, to ensure that it satisfies the business requirements . Approval testing ensures that the item actually solves the customer's problems. It can also be defined as demonstrating that the object meets its intended use when shipped in good condition.

The framework has been successfully attempted and implemented, ensuring that all of the prerequisites specified in the product necessities determination are met.

7.4 Test Cases

Experiments are a set of actions, conditions, and data sources that can be used in the course of executing testing tasks. The main goal of this action is to determine whether a product passes or fails in terms of utility and other perspectives. The process of creating tests can also aid in the discovery of flaws in an application's requirements or design. Experiment serves as the starting point for test execution, and after applying a set of information values, the application produces a definitive result and exits the framework at a specific end point, also known as the execution post condition.

Table 1 Test Cases

S.NO	Test Cases	Pass/Fail
1	Import the necessary libraries and the diabetes dataset.	**Pass**
2	Data should be pre-processed to remove any missing information.	**Pass**
3	Divide the dataset in half, where the Training set receives 80% of the votes, while the	**Pass**

	Test set receives 20%..	
4	Choose between Support Vector Machine, Decision Tree, and Random Forest as your machine learning method.	**Pass**
5	Build a classifier model for the aforementioned machine learning algorithm based on the training set and test the classifier model based on the test set for the aforementioned machine learning method.	**Pass**
6	Make a comparison The outcomes of each classifier's experimental performance are evaluated.	**Pass**
7	After assessing numerous metrics, the highest performing algorithm is determined.	**Pass**

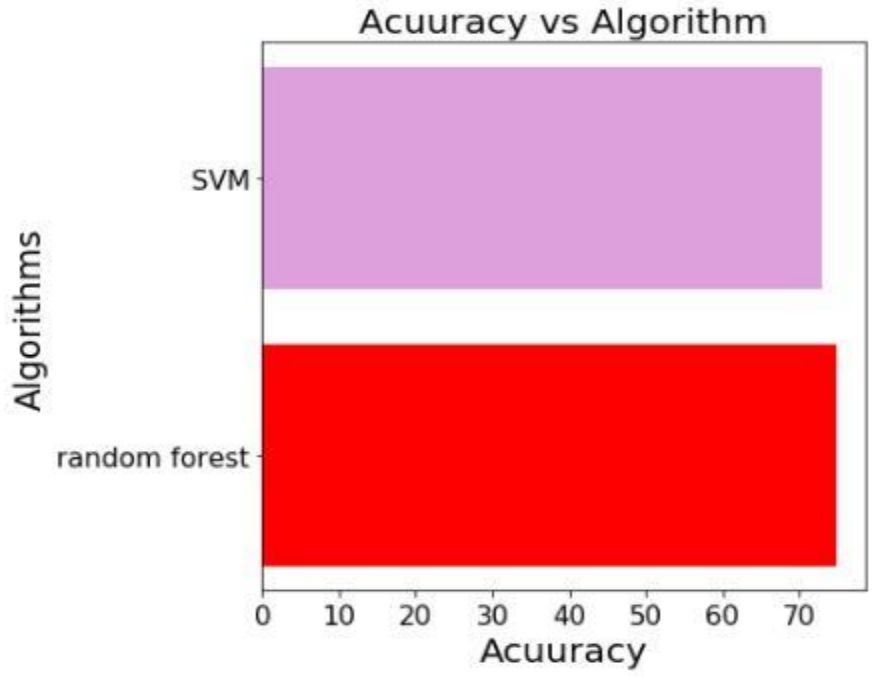
8. RESULTS ANALYSIS

```
data = {'Algorithms':['Random Forest Classifier', 'SVM'],'Accuracy':[accuracy_ranfor,accuracy_svc]}
df = pd.DataFrame(data)
print(df)
```

```
                 Algorithms  Accuracy
0  Random Forest Classifier  0.759740
1                       SVM  0.733766
```

Output 1(screen)

```
abc=[75,73]
plt.figure(figsize=(6,6))
colors=['red','plum']
labels=['random forest','SVM']
plt.barh(labels,abc,color=colors)
plt.xlabel("Acuuracy",fontsize=20)
plt.ylabel("Algorithms",fontsize=20)
plt.title("Acuuracy vs Algorithm",fontsize=20)
plt.xticks(rotation=1,fontsize=15)
plt.yticks(fontsize=15)
plt.show()
```



Output 2(screen)

We can see that random forest is more accurate for our project than support vector machine. For better understanding check the output 2 for visualizing the accuracy using graph.

9.Conclusion & Future Enhancement

9.1 Conclusion

Glucose intolerance in pregnant women is known as GDM (gestational diabetes mellitus). We tested our project on a diabetic patient dataset to see if it was accurate or not. When compared to other models, the accuracy of each model varies. The project effort results in a model that is accurate or higher in accuracy, proving that the model can accurately predict diabetes. Our findings reveal that Random Forest outperformed other machine learning algorithms in terms of accuracy. In medical data, missing values are fairly common. Missing value handling is a difficult but fascinating subject of research in the context of information retrieval. We've also dealt with missing values.

9.2 Future Enhancement

In the future, we may be able to recommend meals to each individual in order to control their daily meal intake and maintain a healthy bmi, glucose, and insulin level. This will aid diabetic patients in maintaining control. Furthermore, we can supply meals to individuals who are predisposed to diseases, with meals tailored to their physical condition prior to the onset of the sickness.

REFERENCES

[1] M. R. Malim et al., “A Markov Chain Model for Diabetes Mellitus Patients,” in Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017), Singapore: Springer Singapore, 2019, pp. 483-488.

[2] S. Mutalib, S. Abdul-Rahman, and A. Mohamed, “Mining frequent patterns for genetic variants associated to diabetes,” in Proceedings - International Workshop on Database and Expert Systems Applications, DEXA, 2014.

[3] W. D. Aumiller and H. A. Dollahite, “Pathogenesis and management of diabetic foot ulcers,” J. Am. Acad. Physician Assist., vol. 28, no. 5, pp. 28-34, May 2015.

[4] X. M. Zhen, X. Li, and C. Chen, “Longer-term outcomes in offspring of GDM mothers treated with metformin versus insulin,” Diabetes Res. Clin. Pract., vol. 144, pp. 82-92, Oct. 2018.

[5] K. KC, S. Shakya, and H. Zhang, “Gestational Diabetes Mellitus and Macrosomia: A Literature Review,” Ann. Nutr. Metab., vol. 66, no. 2, pp. 14-20, 2015.

[6] “Standards of Medical Care in Diabetes--2014,” Diabetes Care, vol. 37, no. Supplement_1, pp. S14-S80, Jan. 2014.

[7] P. Damm, A. Houshmand-Oeregaard, L. Kelstrup, J. Lauenborg, E. R. Mathiesen, and T. D. Clausen, “Gestational diabetes mellitus and long-term consequences for mother and offspring: a view from Denmark,” Diabetologia, vol. 59, no. 7, pp. 1396-1399, Jul. 2016.

[8] S. Anazawa, “Gestational diabetes mellitus,” Nihon Rinsho., vol. 73, no. 12, 2015.

[9] H. C. et al., “Investigating the perceived benefits, barriers and beliefs towards physical activity in pregnancy among women with gestational diabetes mellitus,” Ir. Med. J., 2017.

[10] W.-J. Kim, M.-S. Shim, and J.-Y. Kim, “Management of Hyperglycemia in Type 2 Diabetes: A Patient-Centered Approach Developed by the American Diabetes Association and the European Association for the Study of Diabetes,” J. Korean Diabetes, 2012.

[11] B. Kaur, R. Quek Yu Chin, S. Camps, and C. J. Henry, “The impact of a low glycaemic index (GI) diet on simultaneous measurements of blood glucose and fat oxidation: A whole body calorimetric study,” J. Clin. Transl. Endocrinol., 2016.

[12] T. P. Markovic et al., “Randomized Controlled Trial Investigating the Effects of a Low-Glycemic Index Diet on Pregnancy Outcomes in Women at High Risk of Gestational Diabetes Mellitus: The GI Baby 3 Study,” Diabetes Care, 2016.

[13] S. Pais, D. Parry, K. Petrova, and J. Rowan, “Acceptance of using an ecosystem of mobile apps for use in diabetes clinic for self-management of gestational diabetes mellitus,” in Studies in Health Technology and Informatics, 2017.

[14] S. Islam, J. Leech, C. C. Y. Lin, and L. Chrostowski, “Peak Blood Glucose Prediction Algorithm Following a Meal Intake,” in 2007 Canadian Conference on Electrical and Computer Engineering, 2007, pp. 579-582.

[15] A. Purwar and S. K. Singh, “Hybrid prediction model with missing value imputation for medical data,” Expert Syst. Appl., vol. 42, no. 13, pp. 5621-5631, Aug. 2015.

[16] K. A. Eskaf, “Blood glucose level prediction for diabetic patients using intelligent techniques,” 2011.

[17] E. Pustozerov, P. Popova, A. Tkachuk, Y. Bolotko, Z. Yuldashev, and E. Grineva, “Development and Evaluation of a Mobile Personalized Blood Glucose Prediction System for Patients With Gestational Diabetes Mellitus,” JMIR mHealth uHealth, vol. 6, no. 1, p. e6, Jan. 2018.

[18] D. Zeevi et al., “Personalized Nutrition by Prediction of Glycemic Responses,” Cell, vol. 163, no. 5, pp. 10791094, Nov. 2015.

[19] N. Nai-arun and R. Moungmai, “Comparison of Classifiers for the Risk of Diabetes Prediction,” Procedia Comput. Sci., vol. 69, pp. 132-142, 2015.

[20] L. Grandinetti and O. Pisacane, “Web based prediction for diabetes treatment,” Futur. Gener. Comput. Syst., vol. 27, no. 2, pp. 139-147, Feb. 2011.

[21] M. P. Sapna, S; Tamilarasi, A; Kumar, “Implementation of Genetic Algorithm in Predicting Diabetes,” Int. J. Comput. Sci. Issues, vol. 9, no. 1, pp. 234-240, 2012.

[22] I. Contreras, J. Vehi, R. Visentin, and M. Vettoretti, “A Hybrid Clustering Prediction for Type 1 Diabetes Aid: Towards Decision Support Systems Based upon Scenario Profile Analysis,” in Proceedings - 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017, 2017.

[23] R. Ali, J. Hussain, M. H. Siddiqi, M. Hussain, and S. Lee, “H2RM: A hybrid rough set reasoning model for prediction and management of diabetes mellitus,” Sensors (Switzerland), 2015.

[24] American Diabetes Association, “National Diabetes Statistics Report, 2014,” Natl. Diabetes Stat. Rep., 2014.

[25] S. S. Ilango, S. Vimal, M. Kaliappan, and P. Subbulakshmi, “Optimization using Artificial Bee Colony based clustering approach for big data,” Cluster Comput., 2019.

[26] F. S. Ismail and N. A. Bakar, “Adaptive mechanism for GA-NN to enhance prediction model,” in Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication - IMCOM ’15, 2015, pp. 1-5.

[27] S. M. Rosli, M. M. Rosli, and R. Nordin, “A mapping study on blood glucose recommender system for patients with gestational diabetes mellitus,” Bull. Electr. Eng. Informatics, 2019.

[28] S. Mirjalili, “Genetic Algorithm,” in Evolutionary Algorithms and Neural Networks: Theory and Applications, Cham: Springer International Publishing, 2019, pp. 43-55.