

# SRS (System / Software Requirements Specification)

Table of Contents
<b>1 Introduction</b>
1.1 Purpose
1.2 Scope
1.3 Product overview
1.3.1 Product perspective
1.3.2 Product functions
1.3.3 User characteristics
1.3.4 Limitations
1.4 Definitions
<b>2. References</b>
<b>3. Specific requirements</b>
3.1 Functions
3.2 Performance requirements
3.3 Usability requirements
3.4 Interface requirements
3.5 Logical database requirements
3.6 Design constraints
3.7 Software system attributes
3.8 Supporting information
<b>4. Verification (parallel to subsections in Section 3)</b>
<b>5. Appendices</b>
5.1 Assumptions and dependencies
5.2 Acronyms and abbreviations

## 1. Introduction

### 1.1 Business purpose

The new system shall estimate if a customer may be at risk of developing a critical condition like diabetes based on customer recollected data. The fact that one in a 100 people with low risk will get diabetes in the next 10 years according to <http://diabetes.org> is important to consider a system that can help to predict if a person is at risk of developing a live treating condition. A Machine learning algorithm can analyze the customer health-related data, and information provided by the users/costumers. This will not only save money for the insurance company but also introduce the information to the customer giving him time to reach out for medical help and improvement on his lifestyle before something worse happens.

This software should be developed because the set of technological solutions will provide benefits for the company, clients, and any user in particular. it will help with the prevention of health issues by being able to predict its likelihood.

### 1.2 Scope

Business domain: [SmartSelectionGroupSPS.com](http://SmartSelectionGroupSPS.com),

The system shall be build base on a random forest which is a meta estimator that uses several decision trees classifiers on various sub-samples of the dataset. Hence the system would be able to make a decision base on information gathered from customer service, sellers, or previous data. The decision made by the system will categorize the customer into groups of healthy, risk, or possible future risk. Giving the reason why it made this decision and showing according to the risk level that makes it classified on those groups.

### 1.3 Product overview

**SmartSelectionGroupSPS** was created in 2022. its main objective is to help companies and customers to find the right insurance plan and preventive care. Based on AI and machine learning algorithm "Random forest classifier" the software can predict customers' needs based on the previously collected information.

#### 1.3.1 Product perspective

The system's relationship to other related products can be defined as a useful element since any system can work without it. however, working with it will bring many monetary benefits for the company, time-saving, good health, and lifestyles awareness to the patient adding to this its accurate prediction in seconds making the process faster and more convenient for companies.

The product can be used as an element of a larger system and in this case, the main requirement will be a dataset with the health history of patients. however, it can also be used by itself. As a requirement will need a dataset with the health history of the patient to be predicted from.

The interface between the product and the larger system would be a floating bottom that once it gets clicked on will generate a prediction bade on the information provided from the data set.

#### 1.3.2 Product functions

Collect health history records for some time. then used the collected records to train itself for accuracy.

Take a decision based on health history, filter the information with a desertion tree and provide feedback in this case place the user/customer in a selected group according to his/her health information.

#### 1.3.3 User characteristics

The target system audience will need a basic level of education. To operate the system is required reading, writing, interpreting skills, and basic computer use knowledge. The system is not implemented for visual impairment individuals since reading is an essential skill to answer any questions when interacting with the system. There is no need for previous experience to operate the system since it is very communicative and explains every step of the process that needs to be done.

#### 1.3.4 Limitations

- Follow HIPPA
- The user shall have an account to use the system.
- The system shall run on android, apple, windows, and Linux devices.
- The system shall be written on phyton.
- The system shall not go down for accuracy maintenance during weekdays.
- The user shall have a reliable internet connection to use the system.

### 1.4 Definitions

A random forest is a meta estimator that fits several decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is controlled with the `max_samples` parameter if `bootstrap=True` (default), otherwise the whole dataset is used to build each tree.

### 2. References

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=random%20forest#sklearn.ensemble.RandomForestClassifier>

<https://towardsdatascience.com/understanding-random-forest-58381e0602d2>

<https://towardsdatascience.com/random-forest-in-python-24d0893d51c0>

<https://abbreviations.yourdictionary.com/articles/insurance-abbreviations-and-acronyms-made-easy.html>

<https://www.analyticsvidhya.com/blog/2022/01/diabetes-prediction-using-machine-learning/>

### 3. Specific Requirements

#### 3.1 Functions

- The login information shall be to components' email/username/or phone number and password.
- The email/username/or phone number shall have a maximum of characters.
- The password shall have a minimum of 8 characters, it must have a number, and a capital letter, and the same letter or number cant be repeated more than 4 times.
- If the customer enters the wrong password more than 5 times the account shall be locked up for 20 minutes.
- The User/Patients or personnel (customer services) shall provide/ fill on the boxes with his personal information: name, last name, email, phone number, and last 4 of SS# for the registration process.
- User/Patients shall provide a valid identification form for the enrollment process.
- The system will use email/text message validation to verify the information.

- The system will test itself for accuracy every month last Sunday from 1 am to 3 pm.
- Last Sunday of the month at 1 am the system will shut down.
- The system will identify the latest data collected.
- The system will perform an accuracy test with the latest data previously identified.
- After the test is done calibration process will start.
- The system will accept, store, and analyze data to make its prediction.
- The system must work properly as expected with an accuracy of 50 and more.
- Personnel, users, etc. will enter the data.
- The system will save this data in its database.
- The system will analyze data using methods like topic modeling.
- The system will make a prediction and display its decision on the screen.
- The system will analyze the data using a decision tree.
- The system will analyze the information entered by the customer.
- The system will use a decision tree to filter the data previously collected.
- The system will generate a health prediction.
- The system will place the client in a specific group according to the prediction.

### 3.2 Performance requirements

*95 % of the transactions shall be processed in less than 1 second.*

The system shall support 100 to 150 simultaneous users.

The system handles support documents, metadata, and database information.

### 3.5 Logical database requirements

#### What problem are you trying to solve?

Many times, it is found that preventing major health issues such as a heart attack or heart failure is much cheaper than treating the health issue itself. By being able to predict the likelihood of a health issue an insurance provider can take measures to prevent it and thus saving the company money and the patient's life

#### What's the motivation for solving this problem?

Reduction of claims payouts and positively affecting lives of clients

#### What would a highly successful solution do for you?

A machine learning model that can predict the likelihood of a heart attack would then allow for immediate flagging of the patient electronically and then in which staff can make contact to allow for intervention to prevent the heart attack.

#### What's a successful solution worth?

In the US it is estimated heart attacks cost 320 billion a year, even only if 10% of heart attacks were stopped by the machine learning model this would offer 32000000000-dollar savings to the US as a whole

#### Who could influence this project?

The customers, managers, project owners, and personnel.

#### Who could be influenced by this project?

Customers, business owners, employees.

#### Are there any related projects to this one?

No

#### Which activities should be included in the scope?

Some concerns are medical record investigation to determine applicable features, how the data is being stored, discussion with health experts to determine the best features for the model, in-depth discussion with staff and stakeholders to determine how they wish the machine learning model to be used (ran nightly, immediately, where results should be displayed and stored, where to store model). An in-depth discussion with current IT staff on how to best integrate the model into the existing framework.

#### Could there be any unintended consequences of the new system?

No.

#### **Elicit business rules questions**

All data must be tightly contained, and the model must be well protected. The Model must meet HIPAA compliance standards. In addition, the model must be checked to follow widely accepted theory on health patterns. All data fed into the model has to be de-identified.

#### **Elicit user requirements questions**

##### **What goals could this product help you accomplish?**

Alert staff to the high probability so that intervention can be performed

##### **What problems do you expect this product to solve?**

Finding and detecting patients that have a high chance of heart attack allows for intervention before the heart attack thus preventing it and saving the insurance provider a claim and the patient possibly their life

##### **What words would you use to describe the product?**

Machine learning model used to predict chronic illness aiding in preventative medicine

##### **What aspect of the product excites you?**

Not only will this save insurance providers money, but it can be used to prescreen patients leading to prevent heart attacks and thus saving lives and leading to a healthier population.

##### **What aspects are most/least valuable to the users?**

The most valuable to the user is the accuracy of the model and its real-time deployment to give immediate feedback. How the model does it to the user and the data pipeline are still important but not of their concern

#### **Elicit non-functional requirements**

##### **What qualities (e.g., efficiency, security, reliability, etc.) are critical for the specific parts of the product?**

Due to the nature of the Data security is very important. In addition, well testing the model is also very important because wrong answers can be life-threatening. Follow up with making sure the model works with the system and does not consume too many resources is critical

#### **Elicit external interfaces features**

##### **What events must the product respond to?**

Whenever new data is fed into the model must be able to predict the outcome and write out and store the information into a table for immediate user response or automated flagging

##### **Can you describe the environment in which the product will be used?**

If you want to reveal exception conditions...

This model is to be deployed in a cloud environment where it will take new data put into the cloud perform its prediction and then save its information back into another table on the cloud. The preferred language is Python.

#### **Reveal more constraints**

##### **What is most important to you about the product?**

Its accuracy.

##### **How would you judge whether the product is a success?**

Based on percent of accuracy for example It is considered successful if reflects a 60% of accuracy or more.

### How should the product be different from the way things are done now?

Currently, health providers review an individual's medical record and make assessments and judgments. However due to the myriad of possible conditions and the complexity of a human body sometimes the onset of a heart attack may be missed. The Machine Learning model offers a consistent and unbiased approach to checking the likelihood of a heart attack and allows the insurance provider to prescreen patients allowing for intervention before the actual event/heart attack.

### Is there anything else we should be asking you.?

It is to be understood that while implementation is key to this project a fair amount of research will be needed to create the actual model. In addition, no level of accuracy can be initially guaranteed this will need to be an agile process where the model will likely need to be adjusted and tuned through multiple iterations to arrive at the best model. In addition, a framework to check the data going into the model needs to be created in case the new data being fed into the model no longer looks like the original data and thus the model might need to be retrained on new data.

## 3.6 Design constraints

Budget under \$10,000.

The software should be done in 4 months.

Every week the software team should show a new implementation that meets the requirements.

The software should have a feature to test itself for accuracy.

The system must follow WCA 2.0 guidelines.

The system must follow Material design guidelines for the android app implementation.

The system must follow HIPPA.

The system must be retrained every month to improve accuracy with new data.

Key	Summary	Description	T	Linked Issues	P	Label
SPS-48	Each button shall have an appropriate description of its functionality.		☐	SPS-12	⬆	Functional, Non-Functional
SPS-47	Each button shall have a different color according to the set of colors chosen by the owner.		☐	SPS-13	⬆	Functional, Non-Functional
SPS-46	The system layout shall have resalting color to make it easier for the user to locate functionality.		☐	SPS-13	⬇	Functional, Non-Functional
SPS-45	The system shall have a button that when click on, initialize the prediction process.		☐	SPS-13	⬆	Functional, Non-Functional
SPS-44	The system shall have an internal database.		☐	SPS-12	⬇	Functional, Non-Functional
SPS-36	The system shall implement a decision tree.		☐	SPS-37	⬆	Functional
SPS-35	The system shall analyze and compare data.		☐	SPS-37	⬆	Functional
SPS-34	The system shall save data by date and time.		☐	SPS-41	⬇	
SPS-33	The system shall access data in the database whenever needed.		☐	SPS-41	⬆	
SPS-32	The system shall store data in database.		☐	SPS-18	⬆	Functional
SPS-31	Stakeholders shall have an internet connection to use the system.		☐	SPS-42	⬆	
SPS-30	The system shall have an internet connection all the time.		☐	SPS-23	⬆	
SPS-29	The system shall display data output.		☐	SPS-38	⬆	
SPS-28	The system shall accept data input.		☐	SPS-17	⬆	
SPS-26	Each month the software shall test its accuracy with the new store data.		☐	SPS-6	⬆	
SPS-25	The software shall follow an accuracy level of more than 50% to be used.		☐	SPS-6	⬆	Non-Functional

## User Cases for SmartSelectionGroupSP

Prepared by Sheimy Paz

SmartSelectionGroup

&lt;2/28/2022

## Use Case List

Primary actors	Use Cases
User/Patients Company Personnel (customer services)	Register Account
Manager, supervisor, Company personnel	Self-test
Manager, supervisor, Company personnel	Entering data into the system
Manager, supervisor, Company personnel	Analyzing information.

## Use Case: Register Account

UC ID and Name:	<b>SSGSP-01</b> Register Account		
Created By:	Sheimy Paz	Date Created:	2/28/2022
Primary Actor:	User/Patients	Secondary Actors:	Company Personnel (customer services)
Trigger:	Click registration button.		
Description:	The User/Patients or personnel (customer services) must provide/ fill on the boxes with his personal information: name, last name, email, phone number, and last 4 of SS# for the registration process		
Preconditions:	<p>The User/Patients or personnel (customer services) must have stable internet access to use the system.</p> <p>The User/Patients or personnel (customer services) must be able to follow writing and reading instructions in English.</p>		
Postconditions:	<p>User/Patients must provide a valid identification form for the enrollment process.</p> <p>The system will use email/text message validation to verify the information.</p>		
Normal Flow:	<p>User/Patients must provide a valid identification form for the enrollment process.</p> <p>The system will use email/text message validation to verify the information.</p>		

Alternative Flows:	SSGSP-01.0.0 User will browse the website. SSGSP-01.1.1 User will click on create a new account. SSGSP-01.2.2 User will fill the blank boxes with the required information. SSGSP-01.3.3 User will receive an email or text message to verify the new account and information. SSGSP-01.4.4 User will not be able to verify the account on its device of option. SSGSP-01.5.5 User will click on the verify account button one more time to receive an email or text message again.
Exceptions:	SSGSP-01.0.0. Users will browse the website. SSGSP-01.1.1.1 User will click on create a new account. SSGSP-01.2.2.2 User will fill the blank boxes with the required information. SSGSP-01.3.3.3 User will receive an email or text message to verify the new account and information. SSGSP-01.4.4.4 User will not be able to verify the account on its device of option. SSGSP-01.5.5.5 User will click on the verify account button one more time to receive an email or text message again. SSGSP-01.6.6.6 User will not be able to validate the information on the device for any external condition. SSGSP-01.7.7.7 User will not be able to successfully open the account. SSGSP-01.8.8.8 User will need to fix that external condition and try again.
Priority:	High
Frequency of Use:	Frequently
Assumptions:	If information is validated account will be successfully created.

#### Use Case: Self-test

UC ID and Name:	<b>SSGSP-02</b> Self-test		
Created By:	Sheimy Paz	Date Created:	2/28/2022
Primary Actor:	Manager, supervisor	Secondary Actors:	Company Personnel
Trigger:	Date/hour (last Sunday of every month)		
Description:	The system will text itself for accuracy every month on last Sunday from 1 am to 3 pm.		
Preconditions:	The system must have been working as expected for the last week		
Postconditions:	The system must work properly as expected with an accuracy of 50 and more.		
Normal Flow	Last Sunday of the month at 1 am the The system will shoot down. System will identify the latest data collected. The system will perform an accuracy test with the latest data previously identify. After the test is done calibration process will start.		

Alternative Flows:	SSGSP-02.0.0 Last Sunday of the month at 1 am the system will shoot down. SSGSP-02.1.1 System will identify the latest data collected. SSGSP-02.2.2 System will not identify new data since there was no new data collected. SSGSP-02.3.3 System will perform an accuracy test with the data it already has stored (old data). SSGSP-02.4.4 After the test is done calibration process will start.
Exceptions:	SSGSP-02.0.0.0 Last Sunday of the month at 1 am the system will shoot down. SSGSP-02.1.1.1 System will identify the latest data collected. SSGSP-02.2.2.2 System will perform an accuracy test. SSGSP-02.3.3.3 System accuracy test will be lower than 40%. SSGSP-02.4.4.4 System will redo all processes from step one.
Priority:	High
Frequency of Use:	Frequently
Assumptions:	Test accuracy will be higher than 40%.

#### Use Case: Entering data to the system

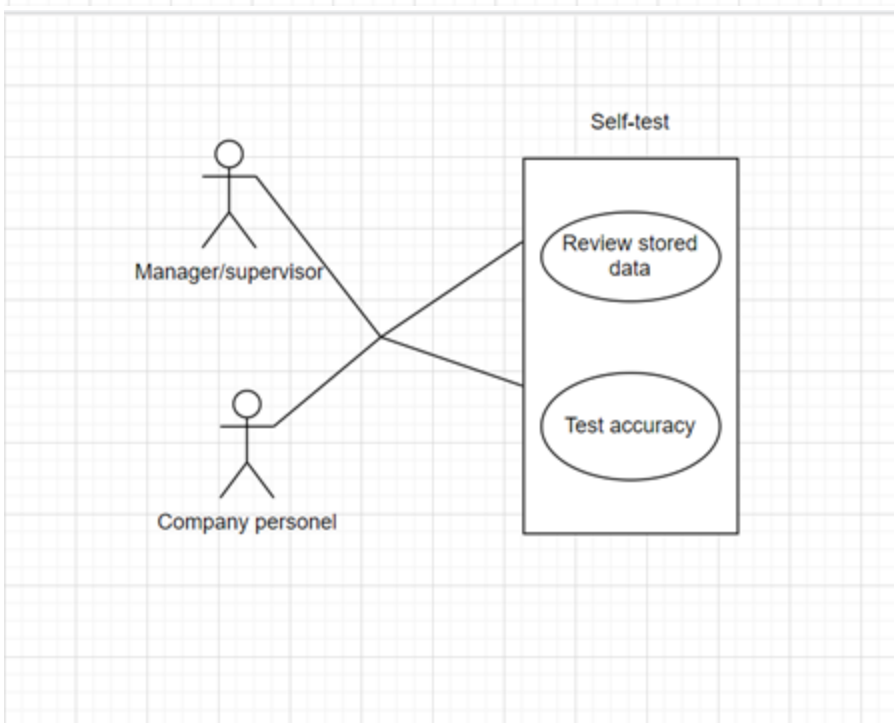
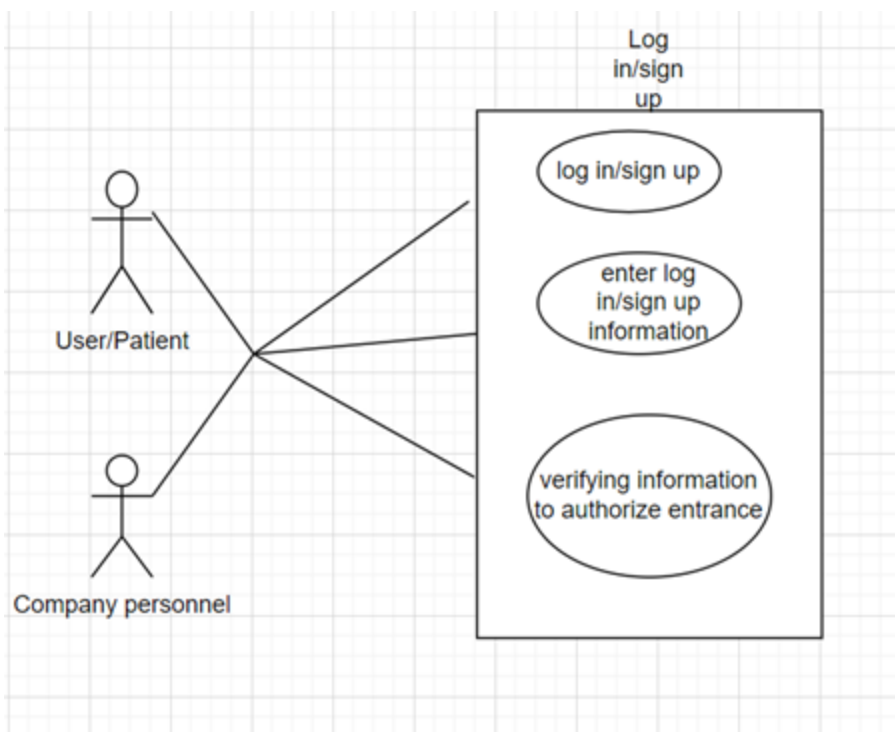
UC ID and Name:	<b>SSGSP-03</b> Entering data to the system		
Created By:	Sheimy Paz	Date Created:	2/28/2022
Primary Actor:	User, manager, supervisor	Secondary Actors:	Company Personnel
Trigger:	Click a button		
Description:	The system will accept, store, and analyze data to make its prediction		
Preconditions:	System must have data previously collected		
Postconditions:	The system must work properly as expected with an accuracy of 50 and more.		
Normal Flow:	SSGSP-03.0 Personnel, users, etc. will enter the data. SSGSP-03.1 The system will save this data in its database. SSGSP-03.2 System will analyze data using methods like topic modeling SSGSP-03.4 System will make a prediction and display its decision on the screen.		
Alternative Flows:	SSGSP-03.0.0 Personnel, users, etc. will update data from a previous patient who is already on the system. SSGSP-03.1.1 The system will save this data in its database. SSGSP-03.2.2 System will analyze data using methods like topic modeling SSGSP-03.4.4 System will make a prediction and display its decision to the screen.		
Exceptions:	SSGSP-03.0.0.0 If the system does not display a prediction the process will repeat one more time.		
Priority:	High		
Frequency of Use:	Relative often		

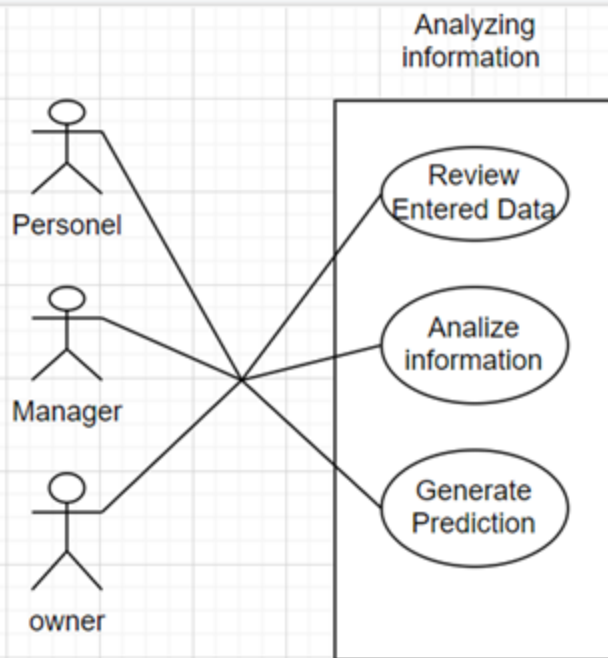
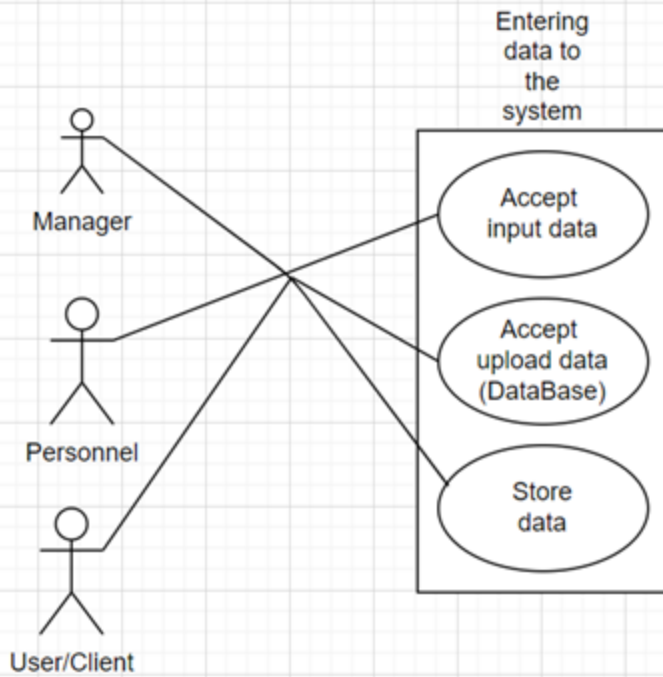



### Use Case: Analyzing information.

UC ID and Name:	<b>SSGSP-04</b> Register Account		
Created By:	Sheimy Paz	Date Created:	2/28/2022
Primary Actor:	Clients, Company Personnel (customer services)	Secondary Actors:	Manager, supervisor, owner
Trigger:	Click the registration button.		
Description:	The system will analyze the data using a decision tree.		
Preconditions:	The System must have the data previously collected		
Postconditions:	The System must work properly as expected with an accuracy of 50 and more.		
Normal Flow:	SSGSP-04.0 The system will analyze the information entered by the customer SSGSP-04.1 System will use a decision tree to filter the data previously collected. SSGSP-04.2 System will generate a health prediction SSGSP-04.3 System will place client in a specific group according to with to the prediction.		
Alternative Flows:	SSGSP-04.0.0 The system will analyze the information entered by the customer SSGSP-04.1.1 System will use a decision tree to filter the data previously collected. SSGSP-04.2.2 System generates a precise health prediction. SSGSP-04.3.3 System will start the process from step one.		
Exceptions:	SSGSP-04.0.0.0 The system will analyze the information entered by the customer SSGSP-04.1.1.1 System will use a decision tree to filter the data previously collected. SSGSP-04.2.2.2 System is not able to generate a precise health prediction. SSGSP-04.3.3.3 System will start the process from step one.		
Priority:	High		
Frequency of Use:	Frequently		

Use case diagrams:





## 1.Verification

Key	Summary	Verification Approach
SPS-48	Each button shall have an appropriate description of its functionality.	

SPS-47	Each button shall have a different color according to the set of colors chosen by the owner.
SPS-46	The system layout shall have resalting color to make it easier for the user to locate functionality.
SPS-45	The system shall have a button that when click on, initialize the prediction process.
SPS-44	The system shall have an internal database.
SPS-36	The system shall implement a decision tree.
SPS-35	The system shall analyze and compare data.
SPS-34	The system shall save data by date and time.
SPS-33	The system shall access data in the database whenever needed.
SPS-32	The system shall store data in database.
SPS-31	Stakeholders shall have an internet connection to use the system.
SPS-30	The system shall have an internet connection all the time.
SPS-29	The system shall display data output.
SPS-28	The system shall accept data input.
SPS-26	Each month the software shall test its accuracy with the new store data.
SPS-25	The software shall follow an accuracy level of more than 50% to be used.

16 issues

## 5. Appendices

### 5.1 Assumptions

The project will increase profitability.

Lives will be saved by its performance.

Client/patients will change their lifestyles after seeing any negative outcome from the evaluation of their health history.

It will help providers by finding early symptoms of the early condition.

It will generate an accurate prediction.

The company will be able to afford 10-20% of new diagnostic cases.

The system will be trained when new information enters into the system for accurate precision.

The system will generate reports in 2 seconds or less after the information has been entered.

### 5.2 Acronyms and abbreviations

- A&G - Administrative and General
- ABN - Advance Beneficiary Notice
- ACT - Assertive Community Treatment
- ADHC - Adult Day Health Care
- ADT - Admission, Discharge, and Transfer
- AIU - Adopt, Implement, and Upgrade
- ALC - Alternate Level of Care
- ALOS - Average Length of Stay