DUY TÂN UNIVERSITY

International School

CAPSTONE PROJECT 2

CMU-SE-451

Architecture Document

v 1.5

GreenBig5 Information System

Submitted by

Loc, Nguyen Tien
Chung, Hoang Bao
Vinh, Do Quang
Kha, Ngo Van

Approved by

Capstone Project 2 - Mentor:

Name	Signature	Date
Nguyen Thanh Binh		15.05.2022

Project acronym	GB5			
Project Title	GreenBig5			
Start Date	01-Mar- 2022	**End Date**		16- May- 2022
Lead Institution	International School, Duy Tan University			
Project Mentor	Doctor. Habil. Binh, Thanh Nguyen			
Scrum master / Project Leader & contact details	Loc, Nguyen Tien Email: nguyentienloc18122000@gmail.com Tel: 0932478789 Student ID: 24211202217			
Partner Organization	N/A			
Project Web URL	N/A			
Team members	Student ID	Name	Email	Tel
	24211207051	Chung, Hoang Bao	Baochunga1@gmail.com	0975059021
	24211208568	Vinh, Do Quang	doquangvinh0708co@gmail.com	0937519169
	24211210573	Kha, Ngo Van	ngokha437@gmail.com	0935950384

DOCUMENT INFORMATION			
Document Title	Architecture Document		
Author(s)	Team C2SE.32		
Role	[GB5] Architecture_v1.5		
Date	25 -Feb- 2022	File name	[GB5] Architecture_v1.5
URL	N/A		
Access	Project and CMU Program		

REVISION HISTORY

Version	Person(s)	Date	Description	Approval
Draft	Kha	25 – Feb - 2022	Initiate document	x
1.0	All members	4 - April – 2022	Finish content of document	x
1.1	All members	6 – April - 2022	Update content	x
1.1.1	Kha, Chung	7 - April - 2022	Add System Context, Container Diagram,	x
1.1.2	Kha, Chung	10 - April - 2022	Add Component, Class Diagram	x
1.1.3	Vinh, Loc	12 - April – 2022	Add Quality Attributes	x
1.2	All members	15 - April - 2022	Update Container Diagram	x
1.3	Chung	17 - April - 2022	Add Allocation Diagram	x
1.4	Kha	20 – April -2002	Update Components Diagram	x
1.5	Chung, Kha, Vinh	16 – May - 2022	Add Location-Based Generate Question, Decision Tree, Update Component Diagram of Single Page	x

TABLE OF CONTENTS

1. INTRODUCTION

1.1. PURPOSE

The purpose of the Architecture document is to:

- Define the architecture needs and technology in detail.
- Provide solutions for business needs.
- Provide overview about resources, schedule, solution and budget for the project.

The architecture merely introduces the project to the student development teams, and provides the up-front information necessary for the team to develop a specification.

1.2. DEFINITIONS, ACRONYMS AND ABBREVIATIONS

1.2.1. Definitions and Acronyms

Acronyms	Definitions
GB5	Green Big 5 Information System
GUI	Graphical User Interface
SDK	Software Development Kit

1.2.2. Diagram Key/Legend

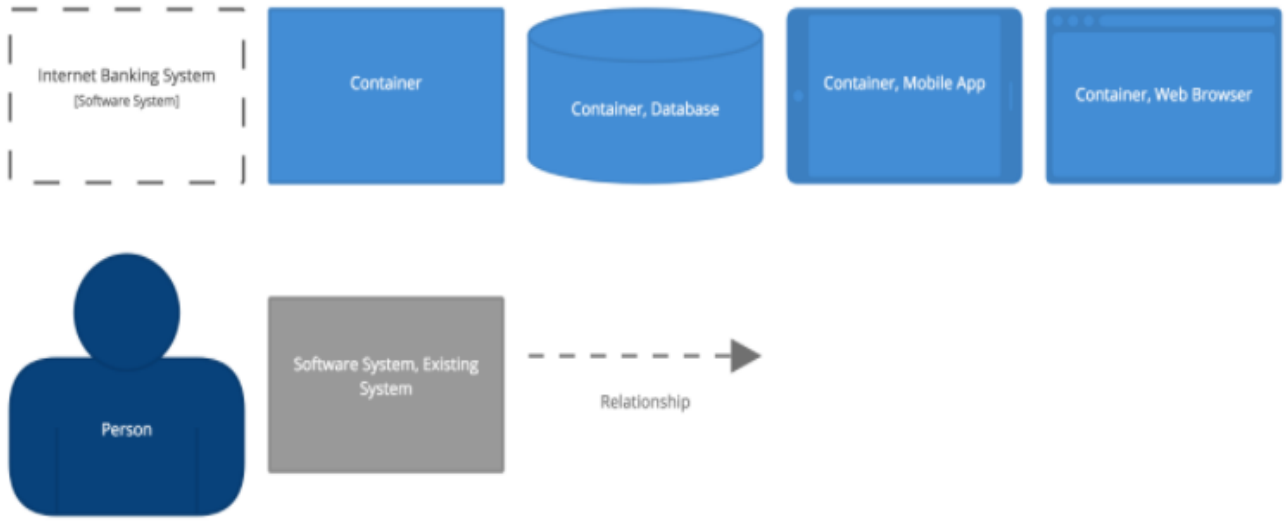


Figure 1.2.2: *Diagram Key/Legend*

1.3. DOCUMENTS REFERENCES

No.	Reference
1	Product Backlog Document for GB5
2	Project Plan Document for GB5

2. PROBLEM STATEMENT

2.1. PROJECT OVERVIEW

As well as the evolution of The Fourth Industrial Revolution (4IR) and the increasing civilization, the environment is becoming harmful by human behavior. Also, at the current rate of urbanization and industrialization, outside of the natural factors, the change of environment is mainly due to human factors. Emissions, population explosion, industrial solid waste, ... are the main causes leading to negative effects on the global environment. To reduce this at a holistic level, predicting human personality and finding the link between it and the environmental impact is the most important task that must be done.

However, predicting human personality and finding the link between it and the environmental impact from many different sources takes a lot of effort and money. To solve this problem, based on our knowledge of big data systems, we have built an intelligent data processing system that can be run on a website-platform with an intuitive and easy-to-use dashboard. This system is a prospective and useful tool for environmental experts and policy makers in Vietnam in particular, and worldwide in general. It can predict user personality and find their effect on the environment and suggest the solution to reduce it.

2.2. BUSINESS DRIVERS

Business problem:

Our environment is always changing. However, at the current rate of urbanization and industrialization, outside of the natural factors, the change of environment is mainly due to human factors. Emissions, population explosion, industrial solid waste, ... are the main causes leading to negative effects on the global environment. To address this at a holistic level, find out the collaboration between human personality and environmental impact is one of the most important missions.

Business need:

Green Big 5 Information System have specific uses:

- Collecting user data

- Predicting user personality trait
- Predict Big5 traits and environmental impacts based on linkings between Big5 traits and environmental ontologies.

All the things above are based on the functionality of the Green Big5 Information System. GB5 fully meets these requirements. Therefore, the development of GB5 is very necessary and meaningful.

2.3. PROJECT GOAL

The aim of this project is to build a GreenBig 5 information system (GB5), i.e., GB5 App, database and GB5 Dashboard:

- GB5 Dashboard: Support for create question packages which used to direct the user follow the environment theme. With each question, users can be distributed by Indicator (2). Finally, by using a prediction method to predict the user's personality traits (3) and predicts linkings between big5 traits and environmental impacts.

  Implemented through 3 steps:

  1. Expert models: expert users use the dashboard to specify the linkings between big traits -facets and environmental keywords structured in tree formats.
  2. Questions are defined and generated based on the expert model (1).
  3. Predict and verify expert model by using answer results of (2), Based on that. we can verify if the expert model is apllied for which group(s) of users.
- GB5 App (Update): GB5 Application receives user activity data by listening to the event, the state, of the system emitted through Intent so that the system can find out the user's location based.

As a result, government authorities, enterprises, as well as users would have an overview of the environment and have a better solution to change user behaviour and to reduce and prevent it from the bad effect.

3. ARCHITECTURE DRIVERS

3.1. HIGH-LEVEL REQUIREMENTS

(Refer to the Product Backlog document for GB5)

3.2. SYSTEM CONTEXT DIAGRAM

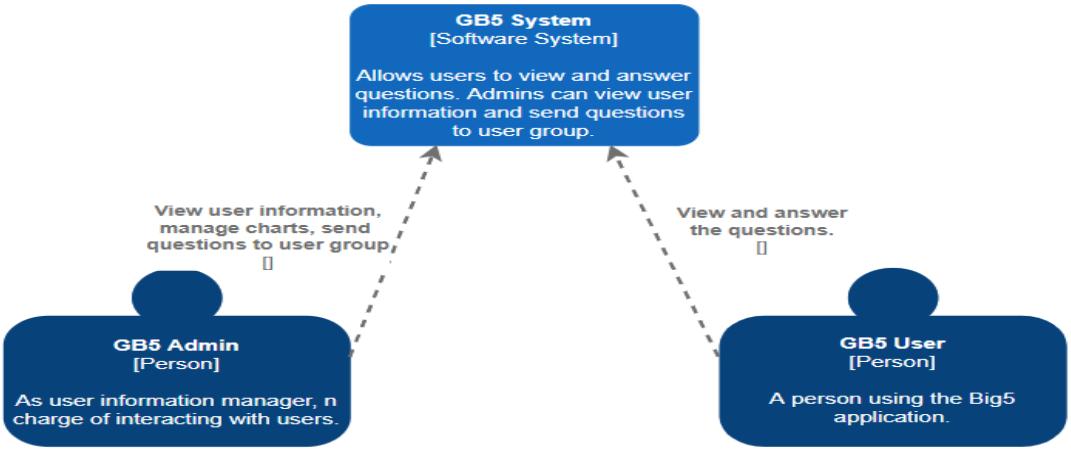


Figure 3.2: Context Diagram of System

3.3. QUALITY ATTRIBUTES

ID	QA01
Quality Attributes	Performance
Stimulus	Submit a question to the user group
Source(s) of stimulus	Admin
Artifacts	System
Environment	Normal mode
System response	The system displays a message that the question has been sent successfully
Response measure(s)	Within 5 seconds

Table 3.3.1: Quality Attributes: Performance

ID	QA02
Quality Attributes	Performance
Stimulus	Login in to the mobile application
Source(s) of stimulus	User
Artifacts	System

Environment	Normal mode
System response	The system displays the question view page
Response measure(s)	Within 3 seconds

Table 3.3.2: Quality Attributes: Performance

ID	QA03
Quality Attributes	Availability
Stimulus	Unable to import question into database
Source(s) of stimulus	Admin
Artifacts	System
Environment	Normal mode
System response	System will log the fault immediately
Response measure(s)	Within immediately

Table 3.3.3: Quality Attributes: Availability

4. CONSTRAINTS

4.1. BUSINESS CONSTRAINTS

- Project will be started on 01 – Mar – 2022
- Project will be finished on 15 - May - 2022
- Duration: 17 weeks

4.2. TECHNICAL CONSTRAINTS

Main Programming Language: Javascripts, Flutter.

- **GB5 Application:**
  - Programming Language: Dart, Flutter.
  - Tool: Android studio.
- **GB5 Database:**
  - Programming Language: NodeJs.
  - Database: MongoDB.
  - Tool: Visual studio code.
- **GB5 Dashboard:**

- Programming Language: NodeJs, ReactJs.
- Tool: Visual studio code.

- **AI Model:**
  - Programming Language: Python.
  - Libraries: sklearn, pandas, pydotplus, fastapi, …
  - Tool: Visual studio code, Pycharm.

5. HIGH-LEVEL ARCHITECTURE

5.1. ARCHITECTURE OVERVIEW

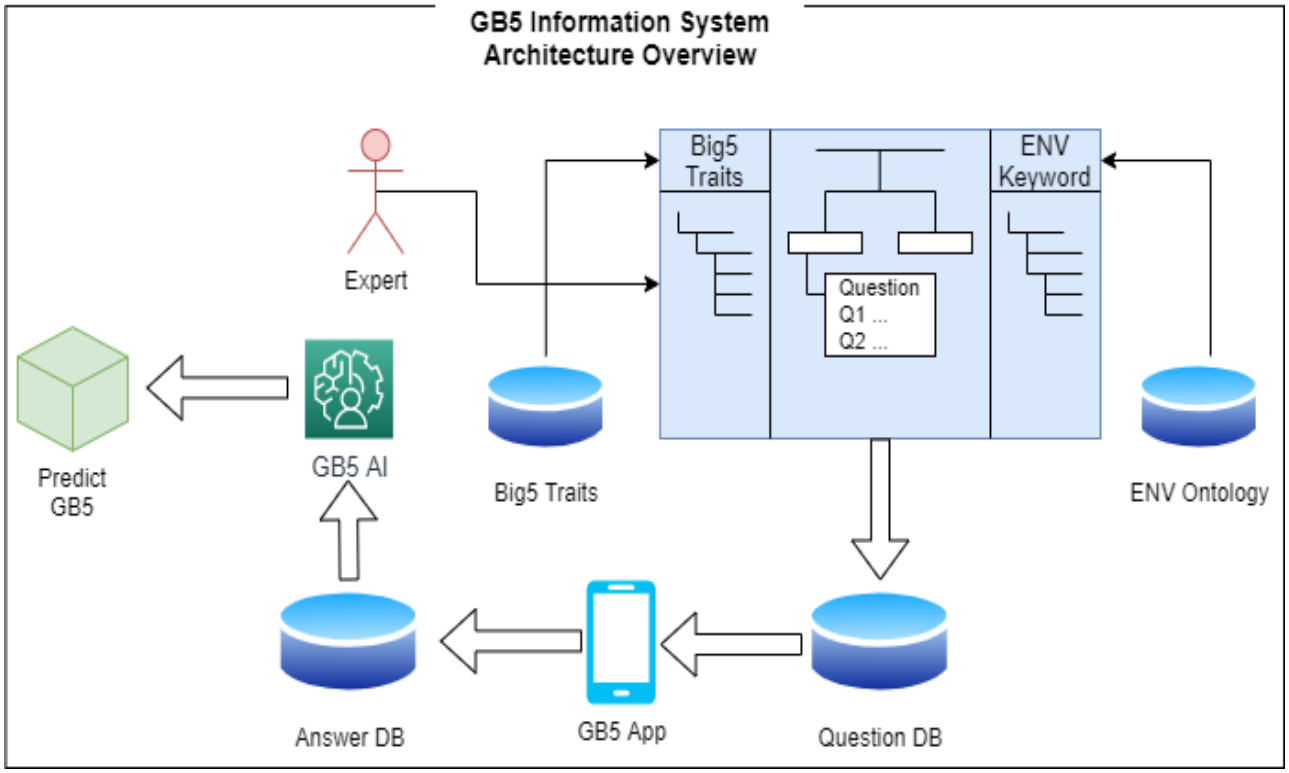


Figure 5.1: *Architecture Overview*

5.2. CONTAINER DIAGRAM

The diagram below shows the overview architecture including containers.

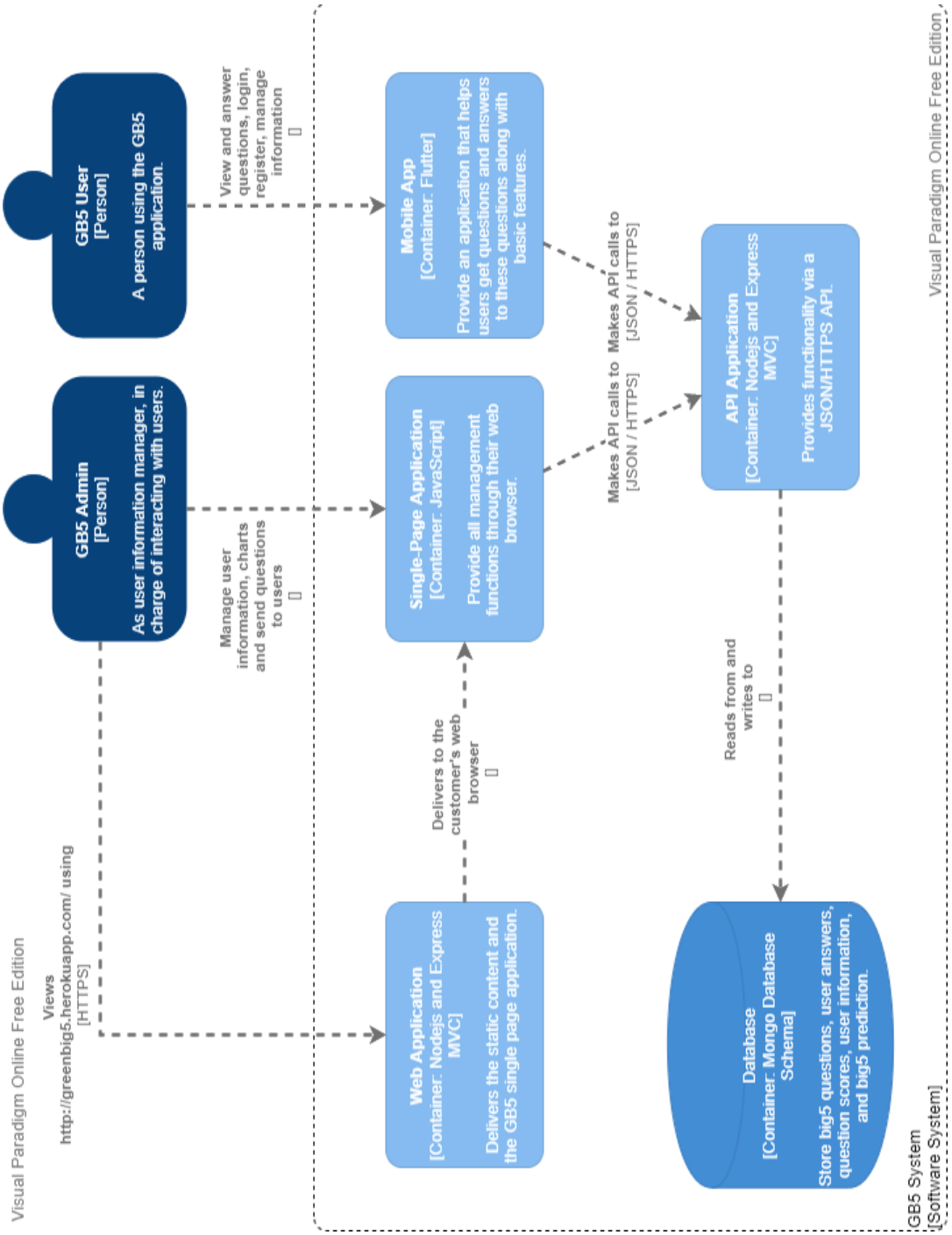


Figure 5.2: *Container Diagram*

5.3. COMPONENT DIAGRAM

5.3.1. Single-Page Application

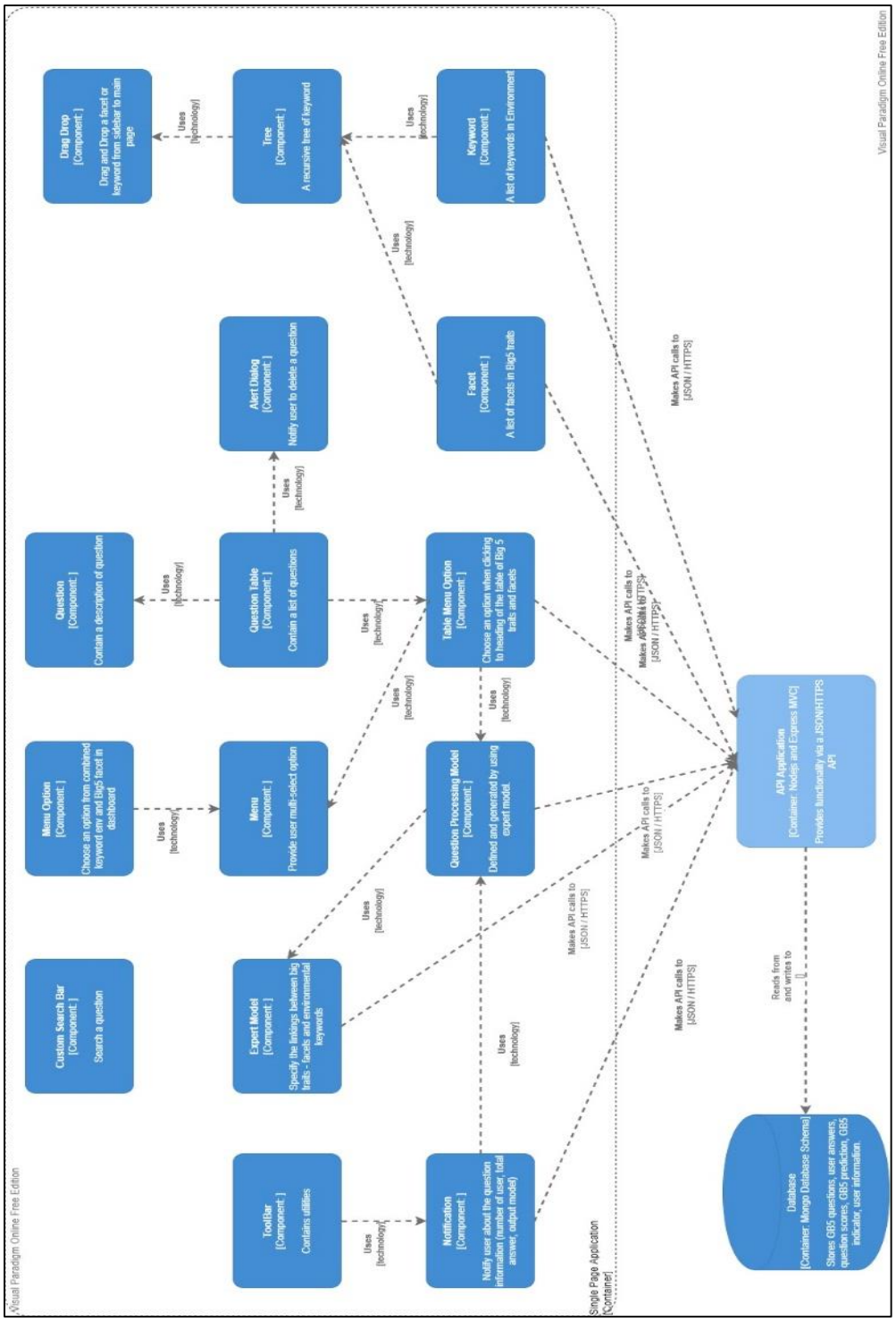


Figure 5.3: *Component diagram of Single-Page Application*

5.3.2. Mobile Application (Update)

The diagram below shows the overview architecture including components and other related components.

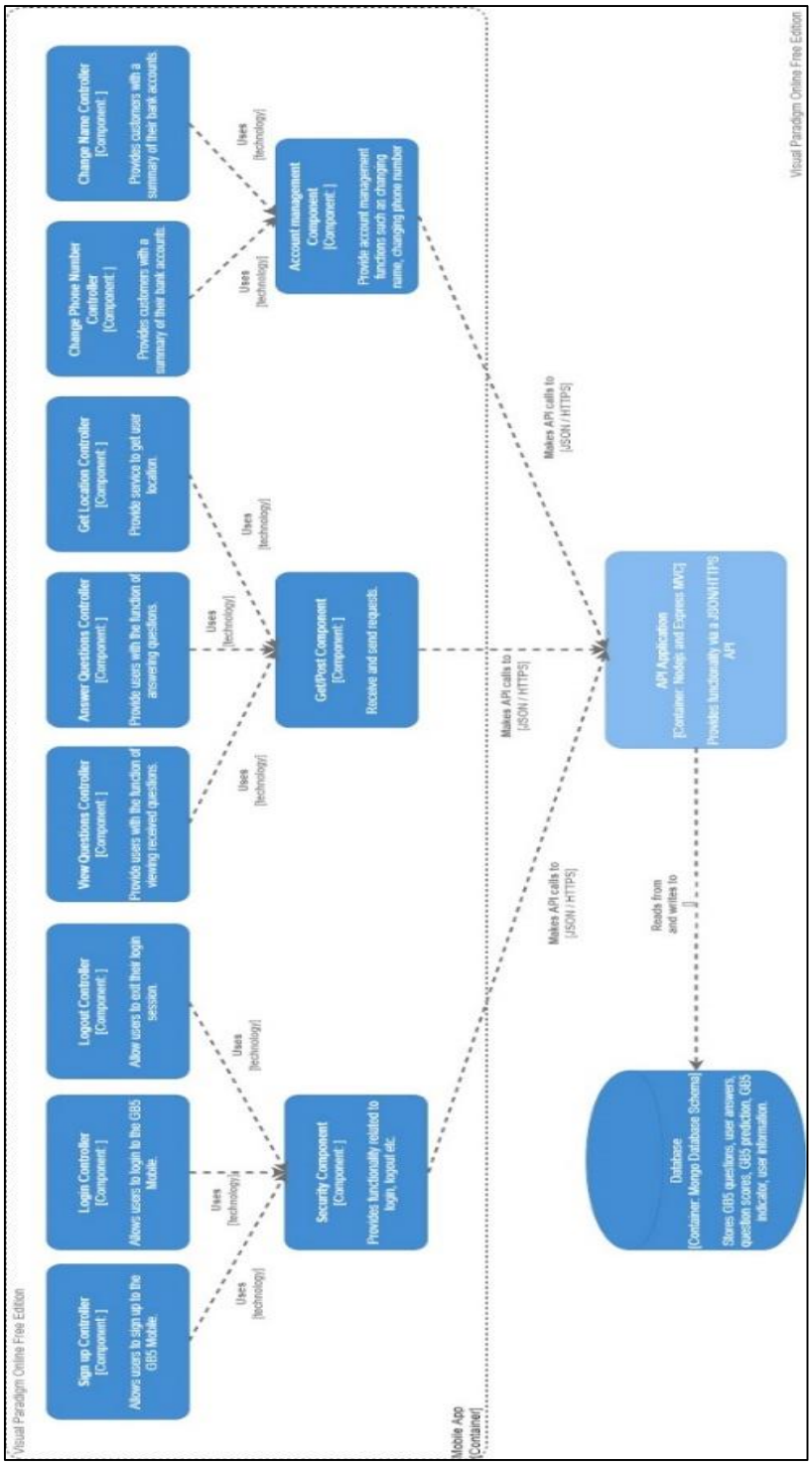


Figure 5.4: *Component diagram of Mobile Application*

5.4. CLASS DIAGRAM

The diagram below shows the overview architecture including the class diagram of the question management component.

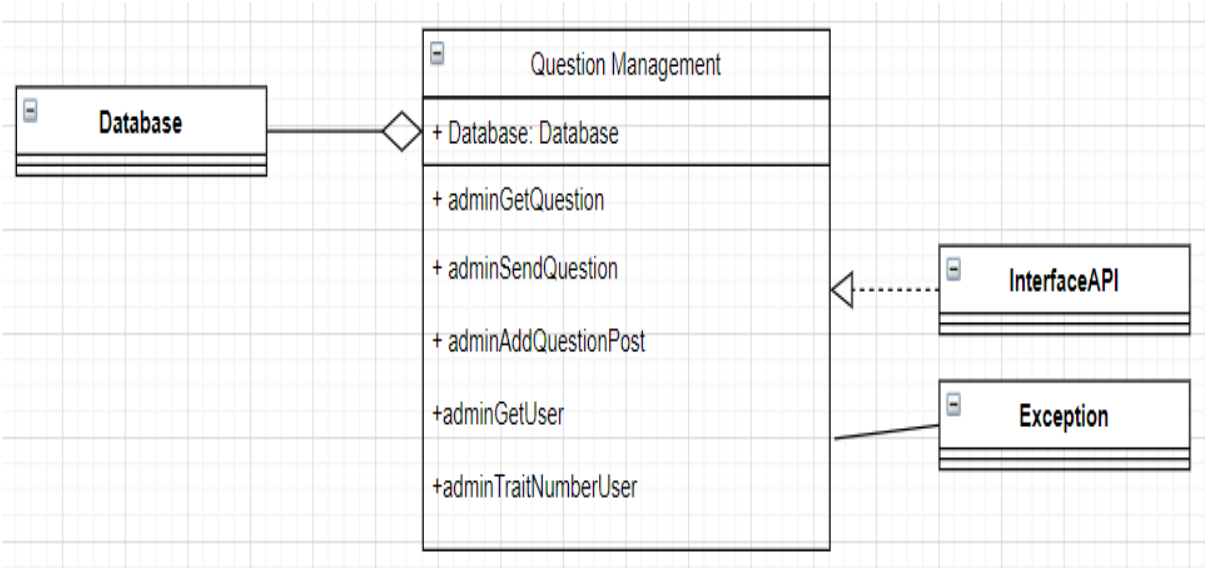


Figure 5.5: *Class Diagram*

Role & Responsibility	Description
adminGetQuestion	Display questions from the database on the board
adminSendQuestion	Submit a question to the user group
adminAddQuestionPost	Enter the question into the database
adminGetUser	Select the User group to submit the question
adminTraitNumberUser	Save indicator information for group classification

5.4 Allocation Diagram

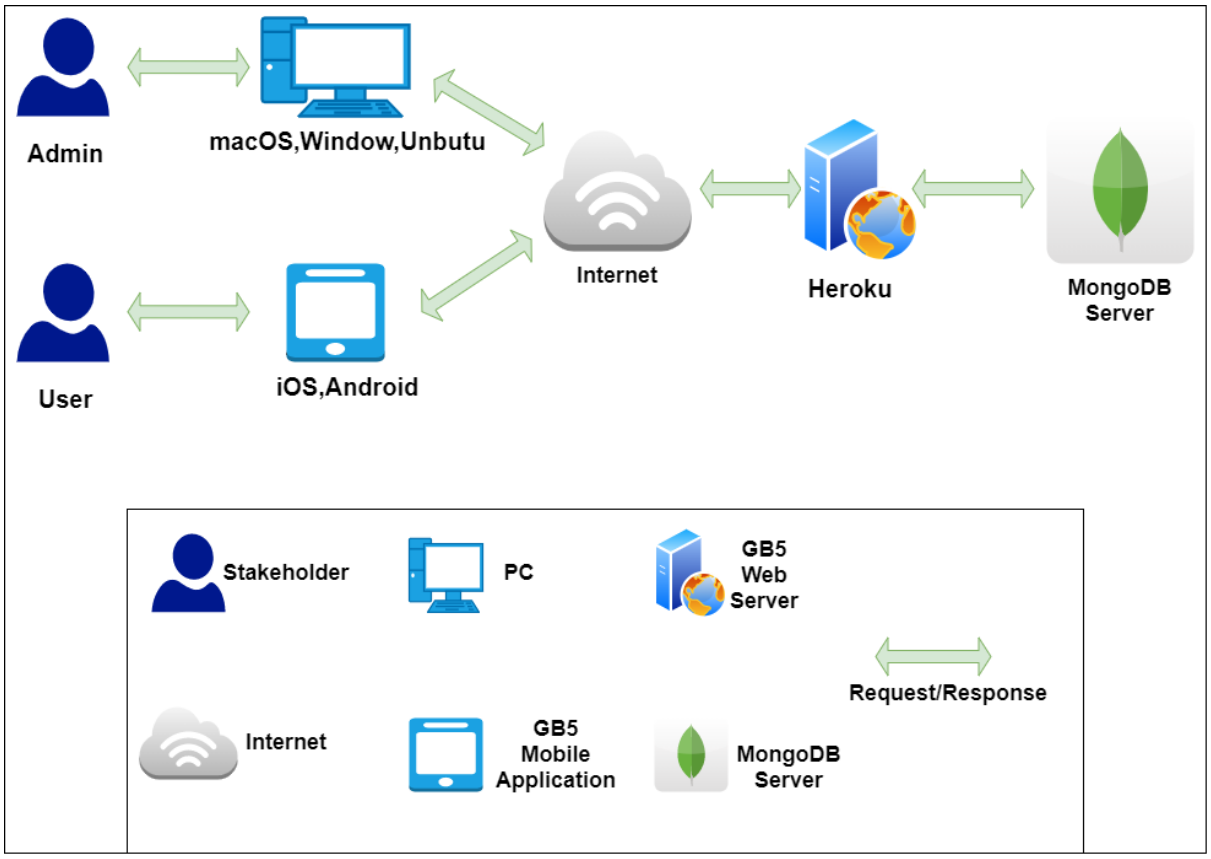


Figure 5.4: *Allocation Diagram*

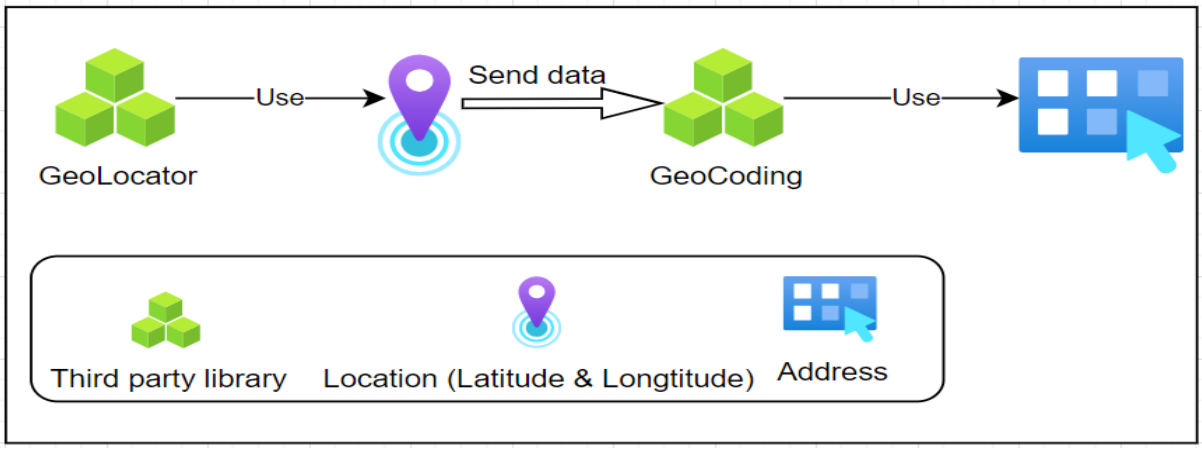
Role & Responsibility	Description
Admin	Admin who interact with GB5 Dashboard Application.
User	User who interact with GB5 Mobile Application.
PC	Devices providing web browsers.
GB5 Mobile Application	The device that provides the operating environment for the GB5 application.
Internet	A global computer network providing a variety of information and communication facilities, consisting of

	interconnected networks using standardized communication protocols.
GB5 Web Server	This is where the API is hosted and provides hosting...
MongoDB	Database server to store data.
Request/Response	Get request data from client and response the data to client.

6. LOW-LEVEL ARCHITECTURE

6.1 Location Based

When the user allows the app to access the location. It is provided by GeoLocator service. The app uses the GeoLocator service to get the current location Latitude and Longitude. Then use GeoCoding service to convert Latitude and Longitude locations received from GeoLocator to addresses which you will have information in details such as street, locality, country ...



These location attributes can be used as features that provide validation of human interactions with the environment.

6.2 Generate Question

First, user need to choose a facet of trait, and a keyword environment. It will automatically combine to the tree and just click the keyword in the tree they want and select it. The table will display and user select Generate Question option. It will take a little bit time. The flow is:

- The Keyword Big5 Environment will send to NodeJS server and sanitize the request payload, and call an API to Python server for handling more complex algorithm
- The sanitized input should be an array like this [Keyword Env, Facet]
- I will extract each word from array and find the synonyms, because I also have a question generation by text or keyword, so we can train and generate more question from text too and push to database. So that the reason why we find synonyms for each keyword and save it in a 'Tags' of each question generated for searching and analyzing the meaning of the question which has been trained from text or keyword before. I did it because I think we'll used it in future.
- After have a list of tag, for each key in input array I generate some related text sentence from it. And combine all text sentence together and from that I generate each question related and push to database and return response to client web

Example:

Keyword: Food

Facet: Ideas

- ⇨ Question: Do you usually plan a party or a trip?
- ⇨ Keyword Big Five Environment: Food Ideas
- ⇨ Tags: [Food, Ideas, Fruit, Apple, Plan]
- ⇨ Point: Point High = 20, Point Medium = 15, Point Low = 10
- ⇨ Personality: Openness = Medium, Conscientious = Low, Extraversion = High, Agreeable = Low, Neuroticism = Low.

6.3 Decision Tree

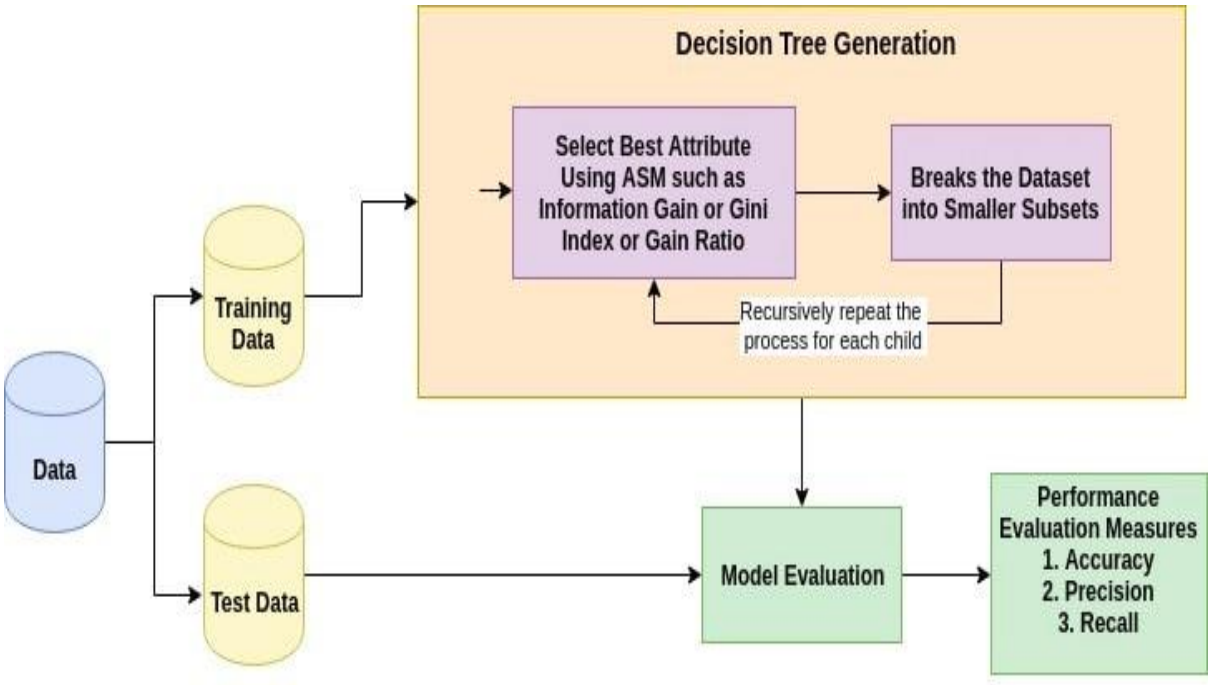
To predict the correctness of the flow generated from the keyword, we use the decision tree algorithm.

- How does the Decision Tree algorithm work?

  The basic idea behind any decision tree algorithm is as follows:

  - Select the best attribute using Attribute Selection Measures (ASM) to split the records.
  - Make that attribute a decision node and breaks the dataset into smaller subsets.
  - Starts tree building by repeating this process recursively for each child until one of the conditions will match:
    - All the tuples belong to the same attribute value.

- There are no more remaining attributes.
- There are no more instances.



- Some formulas need to be calculated in the decision tree:
  - Information Gain

$$Info(D) = -\sum_{i=1}^{m} pi \log_2 pi$$

  - Gain Ratio

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|\mathrm{Dj}|}{|\mathrm{D}|} * \log_2\left(\frac{|\mathrm{Dj}|}{|\mathrm{D}|}\right)$$

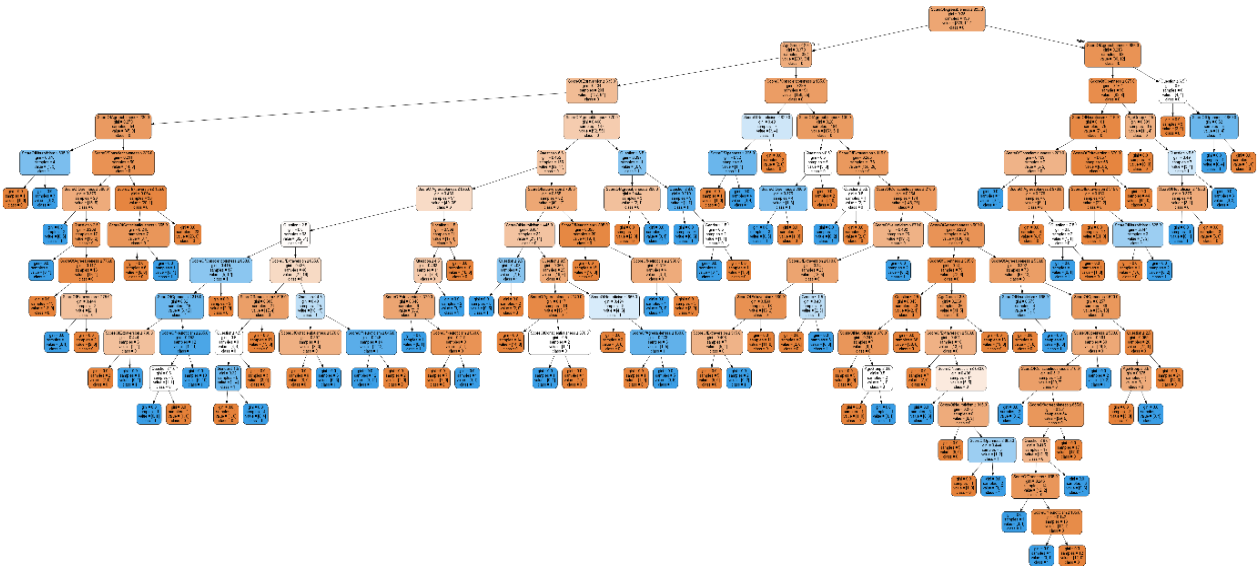
  - Gini index

$$Gini(D) = 1 - \sum_{i=1}^{m} Pi^2$$

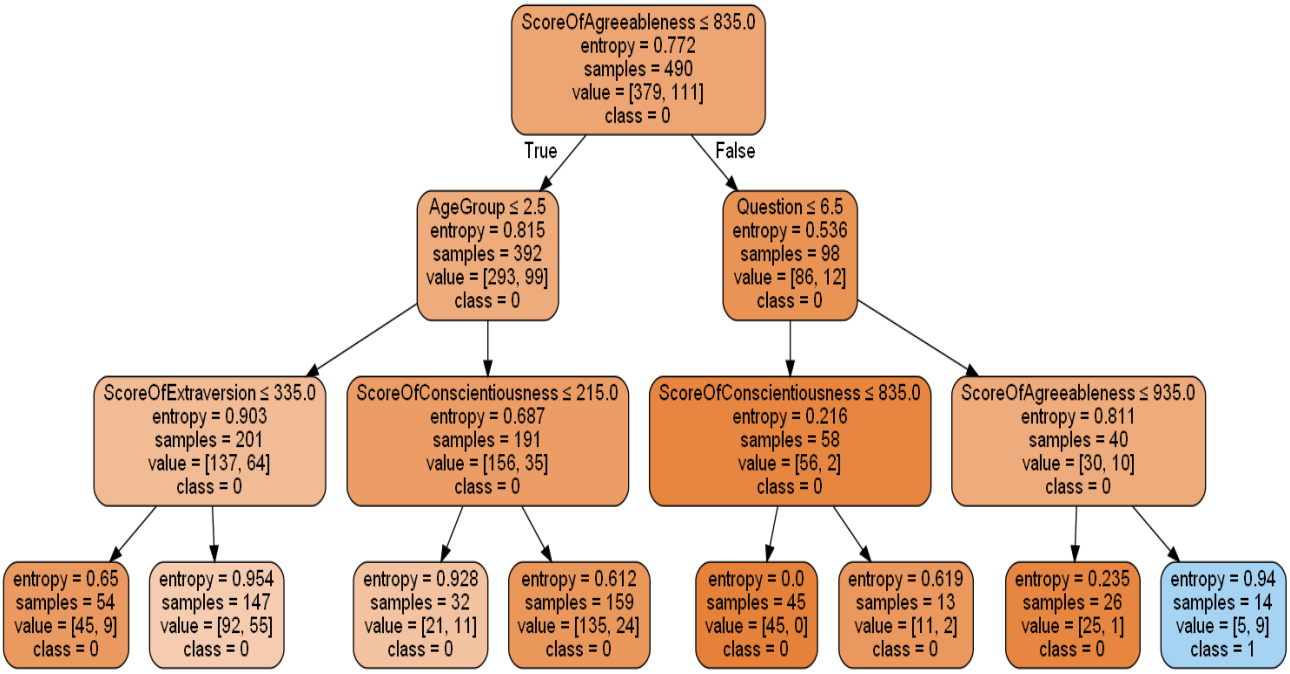
- How do we use this algorithm?
  - The data set after the study is obtained as follows:

AgeGroup	Gender	Location	Question	sO	sC	sE	sA	sN	Outcome
>40	Male	City	Is the rice field the same as a school?	920	470	260	650	110	1
30-40	Female	Country	Is the rice field the same as a school?	440	840	210	820	430	0
<20	Male	Country	Is the rice field the same as a school?	240	480	590	590	400	0
<20	Female	Country	Is there a rice field in texas?	390	850	710	590	420	0
20-30	Female	City	Is there a bad party in the world?	910	370	500	150	550	0
>40	Male	City	Is there a bad party in the world?	930	280	490	190	740	0
30-40	Male	City	Is there a bad party in the world?	720	980	550	370	710	1
30-40	Female	City	Is there a bad party in the world?	970	660	660	560	590	0
>40	Female	City	Is there a bad party in the world?	300	800	680	110	350	1
30-40	Female	Country	Is there a bad party in the world?	730	940	670	380	730	1
<20	Male	Country	Do you like activities about saving electricity?	240	450	820	150	870	1
>40	Female	Country	Do you like activities about saving electricity?	150	210	480	280	910	1
<20	Female	City	Do you like activities about saving electricity?	330	420	660	240	860	1
30-40	Male	Country	Do you like activities about saving electricity?	900	380	560	780	210	0
<20	Female	Country	Do you believe in the fire department to save you when there is a fire?	650	210	830	220	870	0
<20	Male	City	Do you believe in the fire department to save you when there is a fire?	600	810	860	960	180	1
30-40	Male	Country	Do you believe in the fire department to save you when there is a fire?	120	300	370	440	180	1
>40	Male	City	Do you believe in the fire department to save you when there is a fire?	960	490	260	250	270	0
>40	Male	Country	Do you believe in the fire department to save you when there is a fire?	160	110	350	270	180	0
30-40	Male	City	Do you want to use fire	330	860	860	930	830	0
30-40	Male	Country	Do you want to use fire	640	200	610	150	880	0
>40	Male	Country	Do you want to use fire	250	810	360	390	520	0
20-30	Male	Country	Do trees bring value to you?	930	320	660	580	470	1
20-30	Male	Country	Do trees bring value to you?	330	420	710	950	420	0
>40	Male	Country	Do trees bring value to you?	450	210	330	250	410	1
>40	Male	City	Do trees bring value to you?	380	190	150	760	140	0
20-30	Male	City	Do you care about forest fires?	970	140	480	180	330	0
>40	Female	Country	Do you care about forest fires?	940	840	710	1000	680	0

  - Using the decision tree algorithm, we get the following:



- After performance optimization:



➔ From here we can predict whether the newly created flow is true or false with up to 78% accuracy.

7. REFERENCES

- The C4 model for visualising software architecture
- https://online.visual-paradigm.com/
- diagrams.net
- https://protocoderspoint.com/flutter-get-current-location-address-from-latitude-longitude/
- https://codelabs.developers.google.com/codelabs/google-maps-in-flutter#3
- https://www.datacamp.com/tutorial/decision-tree-classification-python