SEMESTER I 2021/2022

TECHNOLOGY AND INFORMATION SYSTEMS

(SECP1513)

SECTION 01

PROJECT 2 - LOW FIDELITY PROTOTYPE

GROUP 2

GROUP MEMBERS:

No.	STUDENT NAME	MATRIC NUMBER
1	AIN BATRISYIA BINTI NORAZLAN	A21EC0009
2	LEE RONG XIAN	A21EC0043
3	MUHAMMAD AKMAL BIN SHAMSUL HAMIDI	A21EC0057
4	NOOR HANNANI SYAMIMI BINTI MOHD SUFFIAN	A21EC0104
5	SITI NURKAMILAH BINTI SAIFUL BAHARI	A21EC0131

CONTENT

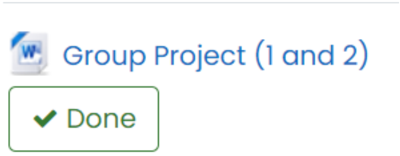
1.0 Introduction

The project is regarding the 4th IR Technology, Machine Learning-Variant Classification. The objective of this project is to develop a low fidelity prototype to solve the problem faced by the client. We have developed a variant classification website to assist bioinformatics scientists and researchers in a clinical reporting workflow. For the first part of the project, there is a detailed description of the client information and an explanation of the Amazon Web Services (AWS) cloud architecture that is used to design the entire ecosystem.

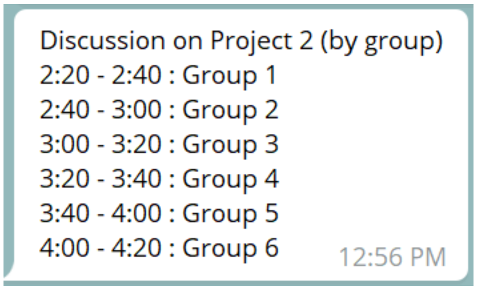
For the second part of the project, the log journal regarding the team progress of the project is table-listed. Moreover, the problems stated by the client which are well-solved with the provided solution as well as team-working are detailed. Besides, the AWS cloud architecture that supports the entire ecosystem has been included as well. Next, the business process flow diagram that involves the client activities in the website has been showcased, followed by the descriptions in a paragraph. Then, a low-fidelity mock-up design using *Marvel* software is displayed. Last but not least, the report is ended by the reflections of team members.

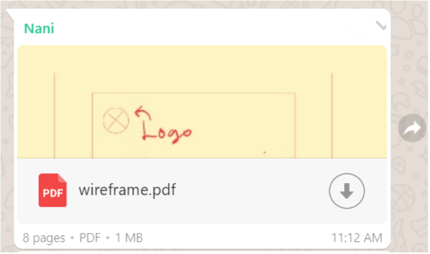
2.0 Log Jurnal

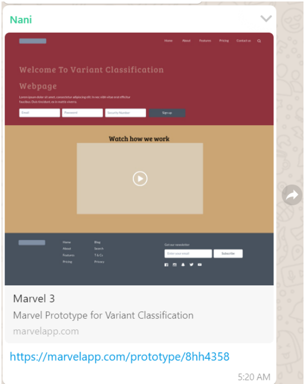
Record of team progress:

No	Date	Team Progress	Remarks
1.	18th October 2021	The students in SECP1513-01 were divided into several groups by Dr Azurah. Each group consists of 5 members. So as to Group 2, the members are: 1. Ain Batrisyia Binti Norazlan 2. Lee Rong Xian 3. Muhamad Akmal Bin Shamsul Hamid 4. Noor Hannani Syamimi Binti Mohd Suffian 5. Siti Nurkamilah Binti Saiful Bahari With that, Group 2 assigned Noor Hannani Syamimi Binti Mohd Suffian as the leader.	
2.	7th November 2021	Dr Azurah introduces the Project-Low Fidelity Prototype to the students of SECP1513 section 01 via E-learning.  The project is divided into part 1 and part 2. This project requires students to design a low fidelity prototype and propose a cloud architecture for the project. The selected project must be based on Fourth Industrial Revolution (IR4.0) technologies such as Artificial Intelligence & Machine Learning, The Internet of Things (IoT), Big Data Analytics, and High-speed mobile Internet.	

3.	20th December 2021	Dr Azurah and the students have a virtual meeting. Dr Azurah urges the students of Bioinformatics to choose the topic regarding Bioinformatics based on Artificial Intelligence & Machine Learning. For project part 1, Dr Azurah urges the students to discuss the case study and solution (using AWS - Cloud Computing) among the group.	
4.	21st December 2021	The members start their action to discuss, everyone does the research for the topic to choose and share their researches through Whatsapp chat group. Eventually, the members decide the topic on variant classification based on Artificial Intelligence & Machine Learning.	
5.	22th December 2021	Group 2 divides the tasks among themselves. For the Project Part 1, 1. Introduction -Siti Nurkamilah Binti Saiful Bahari 2. Client information - Lee Rong Xian 3. Architecture planning and design - Ain Batrisyia Binti Norazlan - Noor Hannani Syamimi Binti Mohd Suffian 4. Conclusion -Akmal Bin Shamsul Hamid	
6.	23th December 2021	The members research the academic articles and journals related to the cloud architecture of Variant Classification. Meanwhile, some of the members refer to the notes of AWS Academy as advised by Dr Azurah. The architecture planning and design are done by referring to the AWS cloud architecture.	

		Some of the members consult about the imaginary client information with Dr Azurah as well.	
7.	26th December 2021	The introduction, client information, Architecture planning and design, conclusion and references are finalised into a document.	
8.	27th December 2021	Submission of Project Part 1 by the leader via E-learning.  SECP1513-01 TEKNOLOGI DAN SISTEM MAKLUMAT (TECHNOLOGY AND INFORMATION SYSTEM) Dashboard My courses SECP1513-01 26 December - 1 January Link for Submission of Project 1 Link for Submission of Project 1 Opened: Saturday, 25 December 2021, 12:00 AM Due: Monday, 27 December 2021, 5:00 PM ✓ Done Link for Submission of Project 1 Submission status Submission status: Submitted for grading Grading status: Not graded Time remaining: Assignment was submitted 2 hours 56 mins early Last modified: Monday, 27 December 2021, 2:03 PM File submissions: TIS PROJECT 1 GROUP 2.pdf 27 December 2021, 2:03 PM Submission comments: Comments (0)	
9.	17th January 2022	Discussion is carried out with Dr Azurah by groups.  Discussion on Project 2 (by group) 2:20 - 2:40 : Group 1 2:40 - 3:00 : Group 2 3:00 - 3:20 : Group 3 3:20 - 3:40 : Group 4 3:40 - 4:00 : Group 5 4:00 - 4:20 : Group 6 12:56 PM As in the diagram above, we enter the virtual meeting by 2.40 pm. The member summarizes to Dr Azurah about Project Part 1. Dr Azurah then guides Group 2 about the details of the prototype. After the meeting, the members start to do research regarding the design of the prototype.	
10.	18th January 2022	The members start to discuss the design of the prototype. The distribution of tasks is divided among the group members. 1. Introduction & AWS Architecture -Siti Nurkamilah Binti Saiful Bahari	

		2. Log Jurnal regarding the team progress -Lee Rong Xian 3. Description about Problems, Solutions and Team Working - Ain Batrisyia Binti Norazlan 4. Business Process Flow Diagram and Description -Akmal Bin Shamsul Hamid 5. Low-Fidelity Mock-Ups - Noor Hannani Syamimi Binti Mohd Suffian Hence, the members start to carry out the completion tasks respectively.	
11.	19th January 2022	One of the members, Hannani, has come up with the wireframe for the mock-up. The members discuss and contribute their ideas for better modification.  Nani wireframe.pdf 8 pages • PDF • 1 MB 11:12 AM Hannani then does it in Marvel, one of the software for the design of mock-ups.	
12.	20th January 2022	Hannani has come up with the prototypes.	

		 Nani Marvel 3 Marvel Prototype for Variant Classification marvelapp.com https://marvelapp.com/prototype/8hh4358 5:20 AM The Log Jurnal has been completed by Lee Rong Xian. The Description about Problems, Solutions and Team Working has been completed by Ain Batrisyia Binti Norazlan. The Business Process Flow Diagram and Description has been completed by Akmal Bin Shamsul Hamid. The Introduction & AWS Architecture has been completed by Siti Nurkamilah Binti Saiful Bahari.	
13.	21st January 2022	The final report has been accomplished and finalized. Then, the members start a discussion with the group video. The main presentor is Siti Nurkamilah Binti Saiful Bahari. The video is then completed.	
14.	22nd January 2022	Final checking is carried out by the group members. After checking, the report together with the video are submitted.	
15.	24th January 2022	Presentation of project group by group via virtual meeting.	

3.0 Problem

There are many challenges we faced while trying to satisfy our client's request to create a machine-learning model based on tertiary analysis to predict if a variant has a conflicting classification. The issue encountered previously is that the researchers take too much time looking for the conflict in the classification of the genomic variant. This variant classification is needed to determine whether a DNA variant causes disease and whether it can impact our health or not, which it relies on multiple lines of evidence. The client demands that the machine-learning model should utilize tertiary analysis to classify variants and achieve high performance in multiple genes and different health conditions.

4.0 Solution

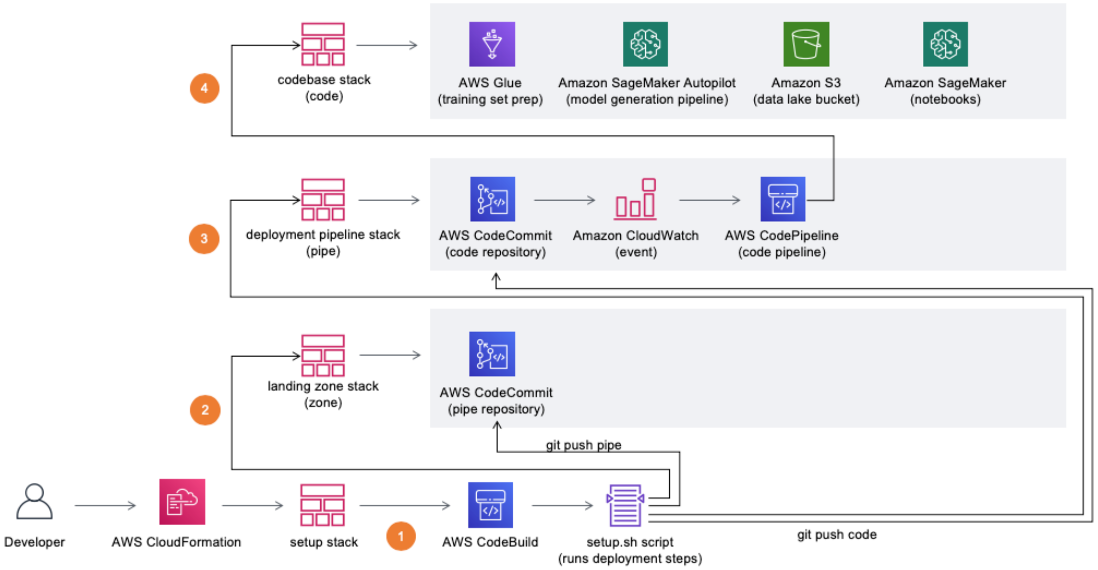
In order to help our client to overcome the conflict in variant classification, we developed a low fidelity machine learning prototype, which is Tertiary Analysis. We use AWS as a platform to help solve our client's request. Why? AWS gives many benefits as it provides security, scalability, adaptability, improved productivity and is also cost-effective to run genomics analysis. We are also confident that AWS could help the client in boosting productivity. All of these benefits makes it easier for us to build a model that can predict the existence of a conflicting categorisation for a variant so that it can help researchers save time by reducing the amount of time they spend searching for such conflicts. We use the services provided by AWS such as Amazon SageMaker, Amazon S3, AWS CodeBuild, AWS CodeCommit, AWS CodePipeline and many more.

5.0 Team Working

Basically, we conduct our discussions and teamwork through Whatsapp group chat. Before dividing tasks, each of us will do as much research as we can and suggest the possible topics in the Whatsapp group chat. Next, we brief about our respective researches and choose the topic among the suggestions. Then, we agree on what needs to be done and by whom. We divide the tasks by voluntarily picking what task we want and taking responsibility to complete the task(s). Then, we planned and complied with meeting schedules and deadlines. We set dates for the completion of each specific task. While completing our own part, we monitored group processes and intervened to correct problems. To conclude, everyone did their part very well and get to fulfill all the requirements of the project without any problems. (As for this project, Kamilah has done for the "Introduction & AWS Architecture" , Rong Xian has done for the "Log Jurnal", Ain has done for the "Description about Problems, Solutions and Team Working", Akmal has done for the "Business Process Flow Diagram" and lastly Hannani has completed "Low-Fidelity Mock-Ups".)

6.0 AWS Architecture Design

The architecture below enables users to deploy automatically using the solution's implementation guide and AWS CloudFormation template.



(Ratan et al., n.d.)

The AWS CloudFormation template builds four CloudFormation stacks in the user's AWS account.

1. A landing zone (zone) stack containing common solution resources and artefacts.

   The CodeCommit pipe repository is created by the landing zone (zone) stack. The setup.sh script publishes source code to the CodeCommit pipe repository after the landing zone (zone) stack has completed its setup.

2. A deployment pipeline (pipe) stack defining the solution's continuous integration and continuous delivery (CI/CD) pipeline.

   The CloudCommit code repository, an Amazon CloudWatch event, and the AWS CodePipeline code pipeline are all created by the deployment pipeline (pipe) stack. The setup.sh script publishes source code to the CodeCommit code repository after the deployment pipeline (pipe) stack is built up.

3. A codebase (code) stack containing ETL scripts, jobs, crawlers, a data catalogue, and notebook resources are among the other stacks.

   The CloudFormation stack is deployed via the CodePipeline (code) pipeline. After the AWS CodePipeline pipelines are set up, Amazon Simple Storage Service (Amazon S3) buckets are created in your account for storing objects, access logs, building artefacts, and data. Source code repositories on CodeCommit. An AWSCodeBuild project for creating code artefacts (for example, third-party data processing libraries). AWS CodePipeline pipeline for automating resource builds and deployment, with AWS Glue jobs and an Amazon SageMaker Jupyter notebook as an example.

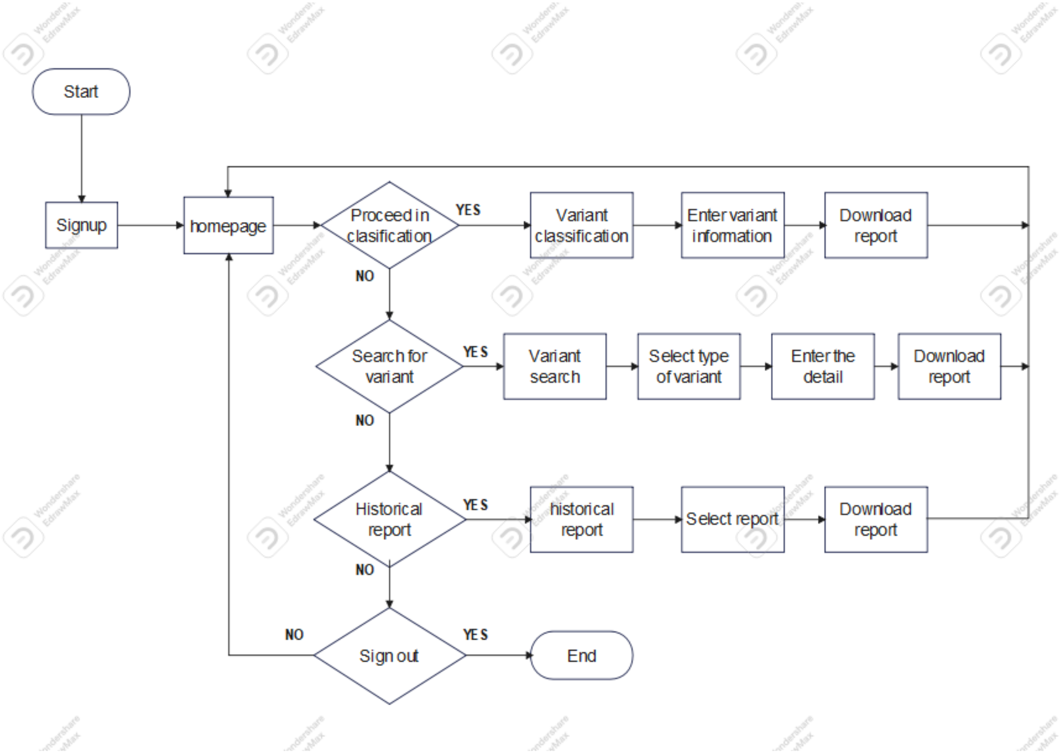
4. Setup stack for installing the solution.

   The setup stack of the solution creates an AWS CodeBuild project with the setup.sh script in it. This script generates the remaining CloudFormation stacks and includes the source code for the AWS CodeCommit pipe repository and the code repository. The example code offers all of the resources needed to create machine learning models and generate genomics data predictions quickly.

Notes:

- A CloudFormation stack is a group of AWS resources managed as a single entity.
- ETL stands for Extract, Transform and Load, and ETL scripts is a script that contains a code that does the extracting, transforming and loading of the data.
- AWS Artifact is the go-to, central repository for all compliance-related information. It gives the user access to AWS' security and compliance reports and several online agreements on-demand.

7.0 Business Process Flow Diagram



Description

Variant classification website is a low-fidelity machine learning prototype. It was developed on purpose to assist bioinformatics scientists and researchers in a clinical reporting workflow. This website also would efficiently classify variants and achieve high performance in multiple genes and different health conditions. The first page of this website is a sign in page. Users have to enter the email, password and security number in order to proceed to the homepage. The homepage consists of three options of the process which are variant classification, variant searcher and record of the past report. The user can download the report of the variant classification when the user enters the variant information. The user also can search for the variant using the variant searcher. Users have to select the type of variant and enter some of the details to download the report of the variant. Users also can view the past report using the past report option on the website by selecting it. When the download of the report is done the website will return to the homepage.

8.0 Low-Fidelity Mock-Up

Link: https://marvelapp.com/prototype/8hh4358

1. **Sign in page or the welcoming page**

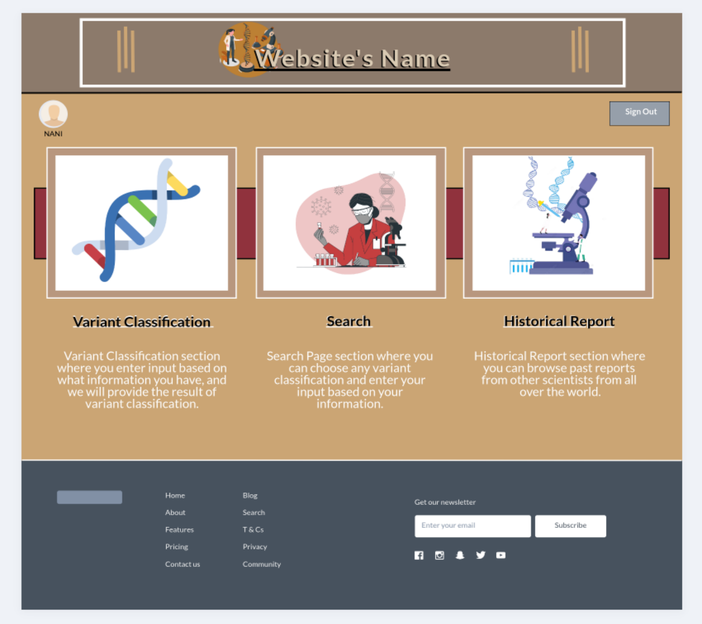
Welcome To Variant Classification Webpage

You can make more confident diagnostic conclusions based on our resultss taken durin our variant categorisation procedure.

Watch how we work

This is a tutorial video to help you navigate throughout our website

2. **Main page**



Variant Classification

Variant Classification section where you enter input based on what information you have, and we will provide the result of variant classification.

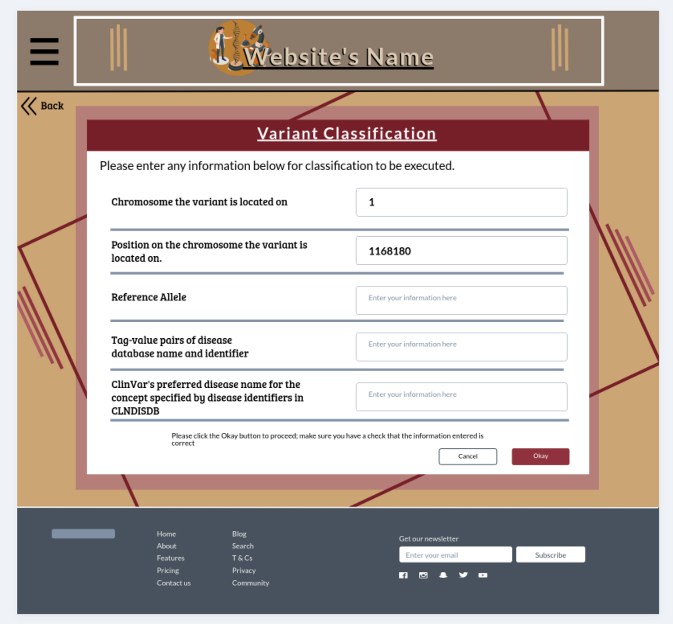
Search

Search Page section where you can choose any variant classification and enter your input based on your information.

Historical Report

Historical Report section where you can browse past reports from other scientists from all over the world.

3. Variant Classification page



Variant Classification

Please enter any information below for classification to be executed.

Chromosome the variant is located on: 1

Position on the chromosome the variant is located on.: 1168180

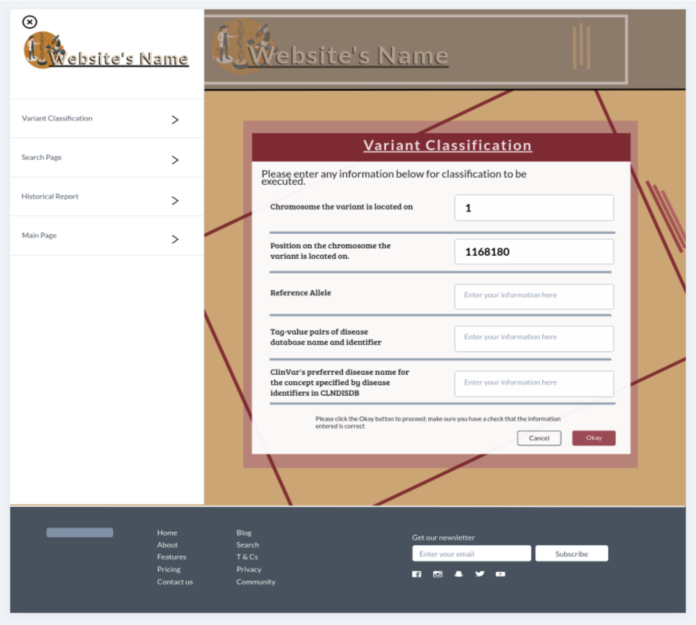
Reference Allele: Enter your information here

Tag-value pairs of disease database name and identifier: Enter your information here

ClinVar's preferred disease name for the concept specified by disease identifiers in CLNDISDB: Enter your information here

Please click the Okay button to proceed; make sure you have a check that the information entered is correct

4. Bar menu



Variant Classification

Please enter any information below for classification to be executed.

Chromosome the variant is located on: 1

Position on the chromosome the variant is located on.: 1168180

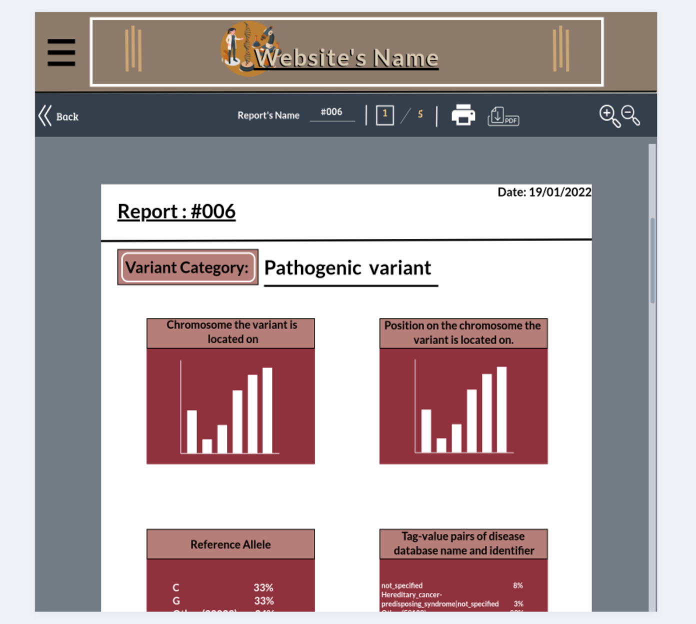
Reference Allele: Enter your information here

Tag-value pairs of disease database name and identifier: Enter your information here

ClinVar's preferred disease name for the concept specified by disease identifiers in CLNDISDB: Enter your information here

Please click the Okay button to proceed; make sure you have a check that the information entered is correct

5. Report page



Date: 19/01/2022

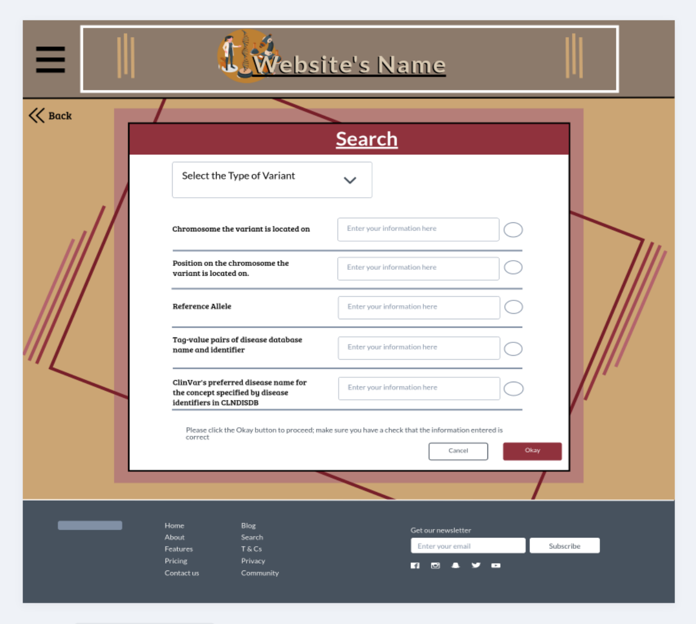
Report : #006

Variant Category: Pathogenic variant

C	33%
G	33%

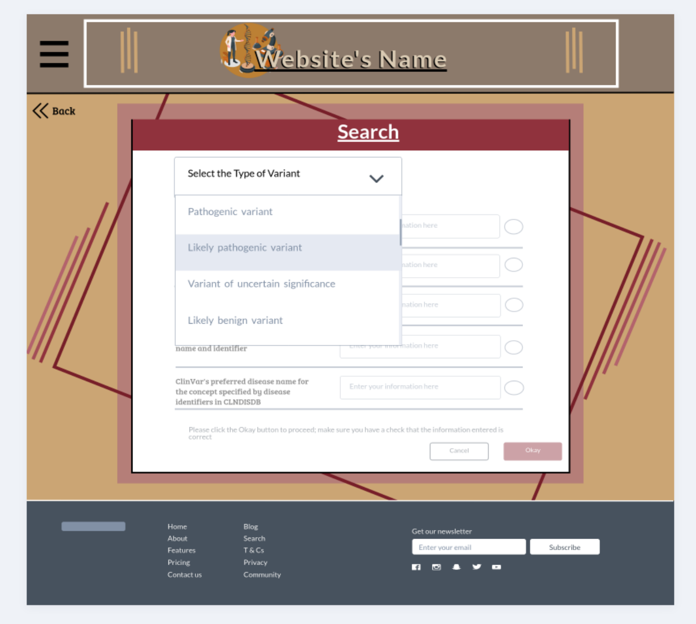
not_specified	8%
Hereditary_cancer-predisposing_syndrome|not_specified	3%

6. Search page



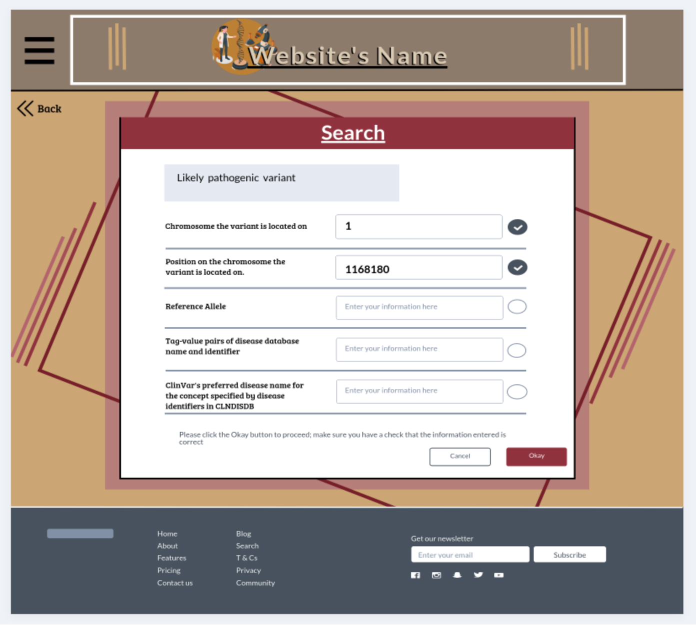
Please click the Okay button to proceed; make sure you have a check that the information entered is correct

7. Search page 2



Please click the Okay button to proceed; make sure you have a check that the information entered is correct

8. Search page 3



Likely pathogenic variant

Chromosome the variant is located on: 1 ✓

Position on the chromosome the variant is located on.: 1168180 ✓

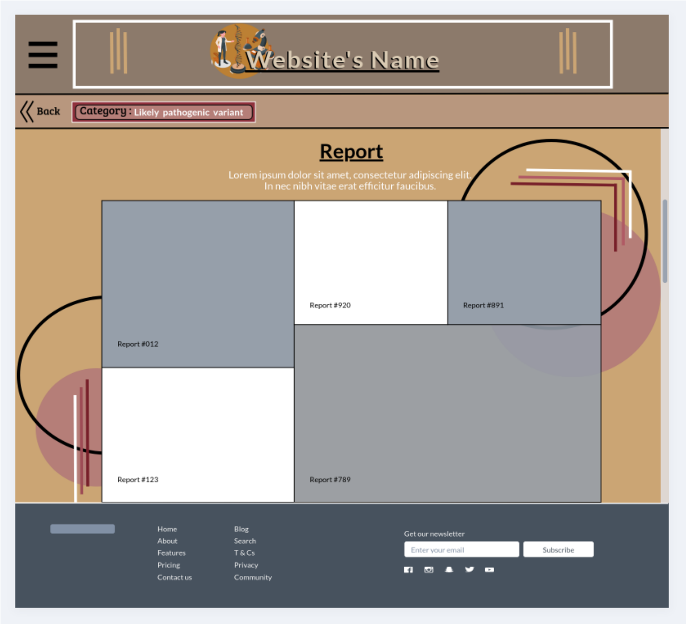
Reference Allele ○

Tag-value pairs of disease database name and identifier ○

ClinVar's preferred disease name for the concept specified by disease identifiers in CLNDISDB ○

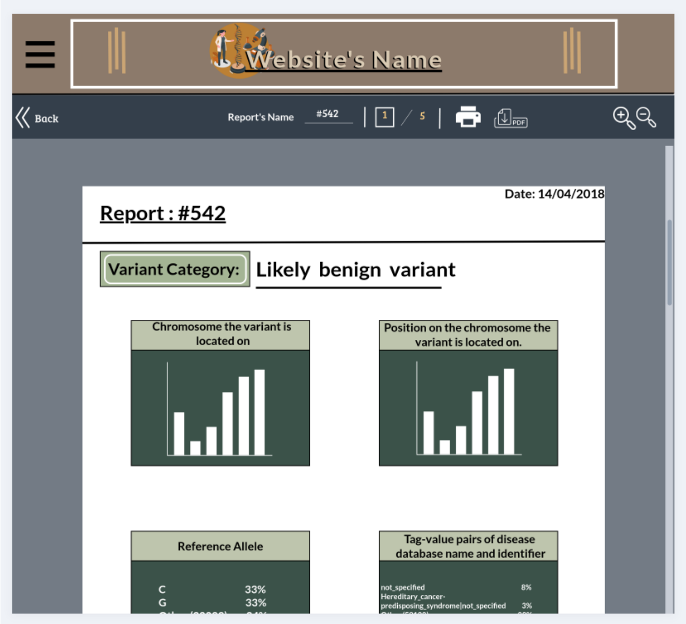
Please click the Okay button to proceed; make sure you have a check that the information entered is correct

9. Search page result



Lorem ipsum dolor sit amet, consectetur adipiscing elit.
In nec nibh vitae erat efficitur faucibus.

10. Report 2



Date: 14/04/2018

Report : #542

Variant Category: Likely benign variant

Reference Allele

C	33%
G	33%

Tag-value pairs of disease database name and identifier

not_specified	8%
Hereditary_cancer-predisposing_syndrome|not_specified	3%

11. Historical report page

12. Report page 3

Date: 08/02/2015

Report : #001

Variant Category: Benign variant

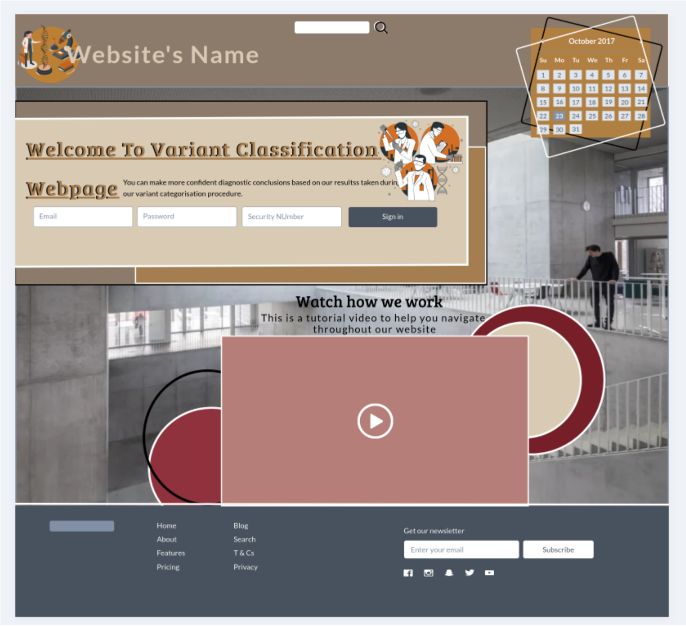
Reference Allele

C	33%
G	33%

Tag-value pairs of disease database name and identifier

not_specified	8%
Hereditary_cancer-predisposing_syndrome|not_specified	3%

13. Sign out page



Welcome To Variant Classification Webpage

You can make more confident diagnostic conclusions based on our resultss taken durin our variant categorisation procedure.

Watch how we work

This is a tutorial video to help you navigate throughout our website

9.0 Student Reflections

9.1 Ain Batrisyia Binti Norazlan

Throughout this project, I got to know how to create a low fidelity prototype which I was never familiar with. This means that I’m able to do something that is beyond my knowledge and at the same time get to learn new things. The things that motivate me to complete this project is about the knowledge itself. I did my best to make sure all the information given is true and follow the criteria. I can come up with creative ideas to improve something or make something new with my team. It was a very challenging experience because we had to create something with our own ideas and creativity.

Even though we successfully completed this project, there are also some problems we encountered causing a slight misunderstanding. One of the problems is we lacked understanding of what needed to be done in this project. We choose one topic for our project which is about variant classification, but since we don’t have much knowledge about the topic, we struggle a lot to do our research. Anyways we managed to complete the assignment, thanks to my teammate that is always there to help.

After going through so much while completing this project, I have found my new interest. Before this, I was very stressed with the Information Technology course because I do not know much and for me, the project given is very difficult to complete with the knowledge I had at that time. Because of this project, I did a lot of research causing me to be able to gather more information and at the same time improve my knowledge about variants.

I believe there are many improvements necessary for me to improve my potential in the industry. Firstly, I will read something new about technology every day to make sure I don’t miss out on current issues. I also need to have analytical and research skills so that I can give out creative ideas on the project I work on. After all, teamwork is the best way to help me to improve my potential in the industry. Working in a group effectively can make all processes smoother.

9.2 Lee Rong Xian

Throughout the project, I have learned the process to develop a Low Fidelity Prototype using AWS architecture and understand roughly how the software developer develops an app or software. Besides, I have realized that Machine Learning is greatly implemented in the field of Bioinformatics. Hence, we have chosen variant classification as our topic among the wide usage of Machine Learning in Bioinformatics. My motivation to complete this project is that it is interesting to build a prototype and the support of my team members throughout the project.

Personally, I think the greatest issue in this project is we lack knowledge in the field of Bioinformatics as our current knowledge is not yet covered as that far and deep. To resolve this issue, I have spent a lot of time doing research on various online resources and platforms such as Google and Youtube regarding the topics as well as discussing with my friends and team members. Moreover, there is too limited time for us to complete the project as there are still several assignments from other subjects that have to be settled. Thus, I have allocated the time correctly for me to settle this project in the meantime, I have also completed other assignments. Lastly, I believe that whenever there is a problem, there must be a solution so we should uphold this belief and determine until success.

Through this first-time experience in developing the low fidelity prototype, prototyping had surprised me which I had never expected before. In the future, I will grab more chances to participate in these similar activities especially to develop a high fidelity prototype to enhance my knowledge and experiences so that I have a greater chance of employability after graduation.

There are a lot of improvements necessary for me to improve my potential in the industry. First of all, I think I have to get explored to the information and technology systems by listening to industrial talk so that I have the knowledge on how our world is transformed. Besides, I need to cultivate and train my thinking skills and creativity to deal with every single problem and to be outstanding in the field of Computer Science.

9.3 Muhammad Akmal Bin Shamsul Hamidi

I have learned many new things by completing this project. The first thing that I learned is how to develop a low fidelity prototype using AWS architecture. In this project, we come out with something that I think is interesting. The main idea was we want to create software that can assist bioinformatics scientists and researchers in a clinical reporting workflow. So we decided to develop Machine learning and Variant Classification as our main topics. My motivation to complete this project is I think I will gain some new knowledge and make me know more about this industry.

By completing this project, we encountered some problems that lead us to misunderstand. The main issue was when we were about to decide the main idea of our project. This is happening because this is our first experience in this sector. We also lack the knowledge to solve the problem. Anyways, with good teamwork, we manage to overcome all the problems and successfully develop the Variant Classification Website.

After completing the project, developing a low fidelity prototype gave me some new knowledge and a bit of surprise that I was not expecting before. Now I know this is not easy as other people said. It needs a piece of knowledge, experience, and a good effort combined in order to develop a good prototype.

I think there are a lot of improvements that I need to improve to succeed in the industry. The first thing is I need to improve myself with some detailed data. It is because I am not good at details. I also have to change the way I am thinking to be a more creative person and a successful person in this industry.

9.4 Noor Hannani Syamimi Binti Mohd Suffian

The main thing I learned throughout the project was how to create a low-fidelity prototype for a website. I find that the process is more complicated than I thought. However, with the help provided by Dr Azurah, team members, and thorough research on the Internet, I was able to complete the making of this prototype. Apart from that, I was also able to learn more about cloud computing through the AWS Academy Cloud Foundation. Here I can implement cloud computing knowledge theoretically on our group project. The following motivation for me to complete this project was my curiosity towards making a website prototype and the support from my team members.

While completing this project, we faced time constraints due to assignments from other subjects. To solve this matter quickly, Rong Xian helped the team organise by setting a reasonable date and time to gather the information and compile it in the report. In addition, we also lacked knowledge on variant classification or, generally, genetic knowledge. To solve the problem, my group had researched and gathered information on this related topic before we properly started the project.

After having a great experience doing this project, I have a spark of hope and motivation to involve myself more in the 4IR technologies. It excites me the most about what is to come in the future. Plus, learning the process of making the low fidelity prototype opened my mind to how the website has been made. It also helps enhance and sharpen my creative skills as I design. All of this would be a handful for me to grow and promote myself to increase my credibility and employability.

The improvements needed to increase my potential in the industry are immense. First, I need to make an effort in making sure that I need to be alert and get hands-on learning rather than theory, mainly because the technology industry has constantly evolved towards better in a short period. Other than that, I also need to make sure my soft skills are better every day that passes. Without the soft skills, I might have a big problem dealing with any related or unrelated to science computers. Communication, teamwork, stress management and time management are the most crucial skills in the industry.

9.5 Siti Nurkamilah Binti Saiful Bahari

I have learned many new things that I never knew before by completing this project. While doing the research about our topic, I get to learn more about the gene variant and its classification. I finally understand what a prototype is. It is a really new thing for me. Of course, the guidance from our lecturer, Dr. Azurah , and support from my group members keeps motivating me to complete this project. As this is a group project, I need to do my best for my part in this project to make it successful.

About issues, I think there are not many issues that we are facing. One of the issues is to do many things related to the project in the given time. So, our solution is to divide the task with our group members so that everyone will contribute, and also we manage to complete the project on the dot. Other than that we also have a bit of a problem in deciding the main idea and flow of the project as we lack knowledge about the tertiary analysis and bioinformatics field itself but with the help of group discussion and deep research we manage to encounter the problem.

After completing this project, I think I would want to discover more about the technology system and how it changes life nowadays. I need to improve and upgrade my knowledge about technology as it will keep evolving time by time. I also need to up-skill my soft skills and hard skills to stay relevant in the industry. To increase my potential in the industry, I should not be left behind. I have to be as advanced as technology.

10.0 Project Video

Link :

https://drive.google.com/file/d/14zzSfAK_i8Z3ebILboefwF3Vx6I1XwYn/view?usp=drivesdk