AI Weather Forecasting

COMP 8037 Major Project Proposal

Jason Zhou-A01005069
11-6-2022

Table of Contents

1. Student Background

My name is Jason Zhou. I am a CST graduate and also a second-year student in the B-TECH Network Security Option. I have a strong interest in cybersecurity and App development. I have built a few apps using Android studio and Google APIs. I am familiar with Java, C++, Python, NodeJS, HTML, SQL and I am always passionate to learn new technologies.

1.1. Education

British Colombia Institute of Technology (BCIT) - Burnaby, BC

- Bachelor of Technology | 2019 – Present |

  o Network Security Application Development Option

- Diploma of Technology | Graduated 2019 |

  o Computer Systems Technology

Simon Fraser University | 2015-2016|

o Computer Science

1.2. Work Experience

BCIT school project

- ***WePerform***

  Developed with: NodeJS, Vue, Firebase

  Description: WePerform is a web app that provides a platform for talents to express themselves and make money. Performers can use this platform to promote themselves or sell lessons, and users can book their desired performers as well.

- ***Daily Fitness***

  Developed with: Android studio, Java

  Description: Daily fitness is an Android app that records users' daily exercises and provides healthy diet recommendations based on the user's BMI. The app uses the phone's built-in sensor to count the user's daily steps. It also has a notebook function to let users record their progress.

- ***Photo Gallery App***

Developed with: Android studio, Java, Google APIs

Description: Photo Gallery App is an Android application. The app allows users to snap pictures and add tags to the pictures. It has convenient search options, users can search for your photos either by tags or by date.

2. Project Description

The project idea is to use Artificial Intelligence to create a webapp that forecasts weather and nature disasters. Weather plays an important role in everyone's life as it can help us to know when it will be rainy and when it will be sunny. Knowing the weather ahead can help us plan our trips or decide our outfits more easily. Meanwhile, nature hazard predictions can help improve transportation safety, help farmers reduce losses from extreme weather and inform people to evacuate beforehand to prevent deaths and injuries. The reason to use machine learning to predict weather and nature catastrophes because it provides faster and more accurate predictions, not only improves weather forecasting but also bring convenience to people.

3. Problem Statement and Background

Weather forecasting plays an important role in day-to-day lives, it can help people decide when to carry an umbrella or choose what outfit to wear. It also improves transportation safety, and it is beneficial to farmers. Currently, the most commonly used weather forecasting technology relies on the combination of computer calculation and human expert analysis. However, this type of forecast is not always accurate. There were several times the weather forecast said it was sunny and it turned out to be rainy, and because of that I got a cold from getting wet. Moreover, inaccurate predictions can lead to more serious consequences. It involves scenarios in which natural catastrophes occur in places where people haven't received proper warning, causing deaths or injuries that could have been prevented.

Therefore, I always wondering why is weather so hard to predict. After researches, I found out that the ability for meteorologists to predict the weather is limited by three factors [1]: the amount of available data; the time available to analyze it; and the complexity of weather events. However, Artificial intelligence can solve all these problems. First, AI is good at dealing with large amount of data and analyzing complex events in a very short amount of time. Even if the amount of data is narrow or partially missing, with an appropriate algorithm, AI model can still find patterns among the data. Therefore, a well-trained machine learning model can provide a more faster and accurate weather prediction than meteorologists.

The project will use Random Forest (R.F.) algorithm as the machine learning model's algorithm. R.F. is good because its capacity to generalize, lower sensitivity to parameter values, and built-in cross-validation. The weather datasets I used will be downloaded from the meteorology websites, and I will split the datasets into train, test and validate sets. More about the algorithm and training methods will

be described in section 6 Methodology. Finally, I will write a webapp interface using python, and wrap the trained AI model into it for display.

4. Scope and Depth

This project's minimum scope is to create a web application that uses AI models to forecast weather. The process of weather forecasting should be done in a reasonably short amount of time after you select the location. The accuracy of short range (1-2hours) temperature and humidity forecasting should be higher than 95%. The accuracy of long range (>7days) temperature and humidity forecasting should be higher than 90%. For nature disaster forecasting (storm and flood), the model should maintain an accuracy above 90%.

Any additional content will not be vital to the project and are therefore stretch goals. These goals include Volcano eruption forecasting and Tsunami forecasting. I will be glad to add any feature that could enhance the security of the application if there is enough time.

5. Test Plan

Functionality Testing

Test Case	Test Description	Tool Used	Expected Result	Pass/Fail
1	To test if the model can-do short-term (in 2 hours) prediction accurately.	-product	The accuracy of the next two hours' temperature and humidity should be higher than 95%.	
2	To test if the model can-do long-term (in 7 days) prediction accurately.	- product	The accuracy of the next seven days' temperature and humidity should be higher than 90%.	
3	To test if the application can work offline	-product	The app should work offline.	
4	To test if the model still forecasts accurately if the	-product	The prediction still maintains high accuracy	

	input data is partially missing.			
5	To test if the model can predict natural disaster accurately	-product	It should be able to forecast nature disasters like storm and flood	

User/ UI Testing

Test Case	Test Description	Tool Used	Expected Result	Pass/Fail
1	To test if all the buttons, lists, and dropdown menus are displayed correctly and work as intended.	-Product	All elements work properly	
2	To test if the users understand how to use our application	-Product	More than 90% of users should know how to use our product.	

6. Methodology

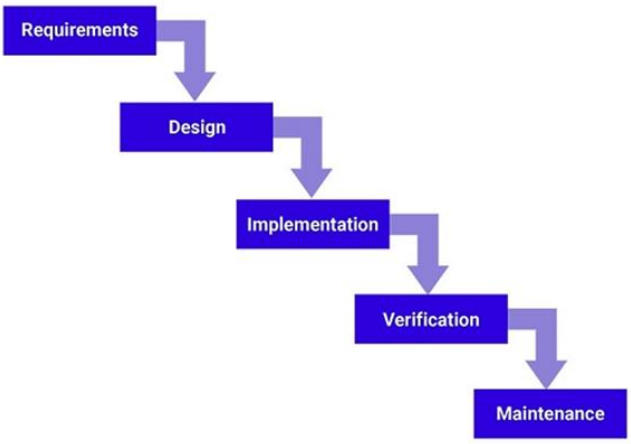
6.1 Technologies and Datasets

The application and machine learning models will be written in Python. The datasets used to train the machine learning model are obtained from the open soured climate data websites/APIs. First, I will combine and clean up the data manually, then use python pandas library to split the data into training set, testing set, and validation set.

6.2 Software development methodology

The project will use Waterfall development method. The waterfall method is a rigid linear model that consists of sequential phases (requirements, design, implementation, verification, maintenance) focusing on distinct goals. Projects with clear objectives and stable requirements can best use the waterfall method. For this project, the objective and requirements are not likely to change and it does not need to update or redesign due to users' feedbacks. Another thing about waterfall method is that each phase must be complete before the next phase can start. For example, I can verify the model only after the training phase is done. Also, I can start training only after the datasets being cleaned and split.

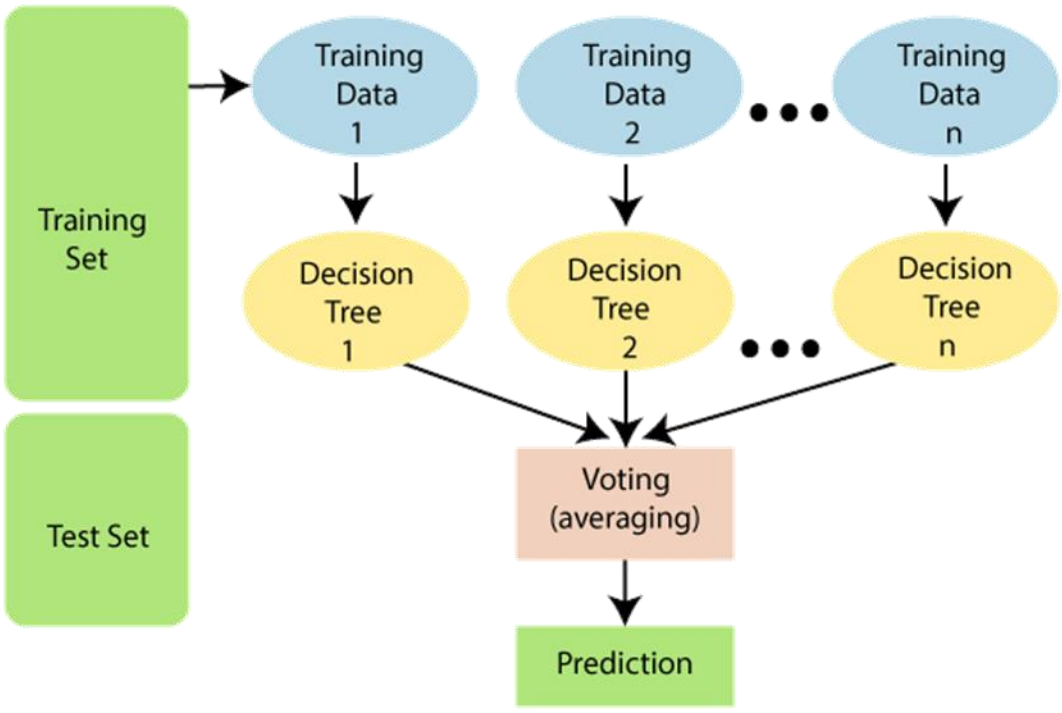
The below diagram describes the waterfall method.



6.3 Algorithms

The algorithm I planned to implement on the model is called the Random Forest (R.F.) algorithm. The R.F. algorithm's major advantages are its capacity to generalize, lower sensitivity to parameter values, and built-in cross-validation. It also has a great speed in processing large amount of data. In simulation long-term monthly temperature, the Random Forest approach is superior compared with other methods. I will use R.F. as my machine learning model's base algorithm, and add more features to continually improves models' accuracy to meet the project need.

The following diagram describes the R.F. algorithm [2].



Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

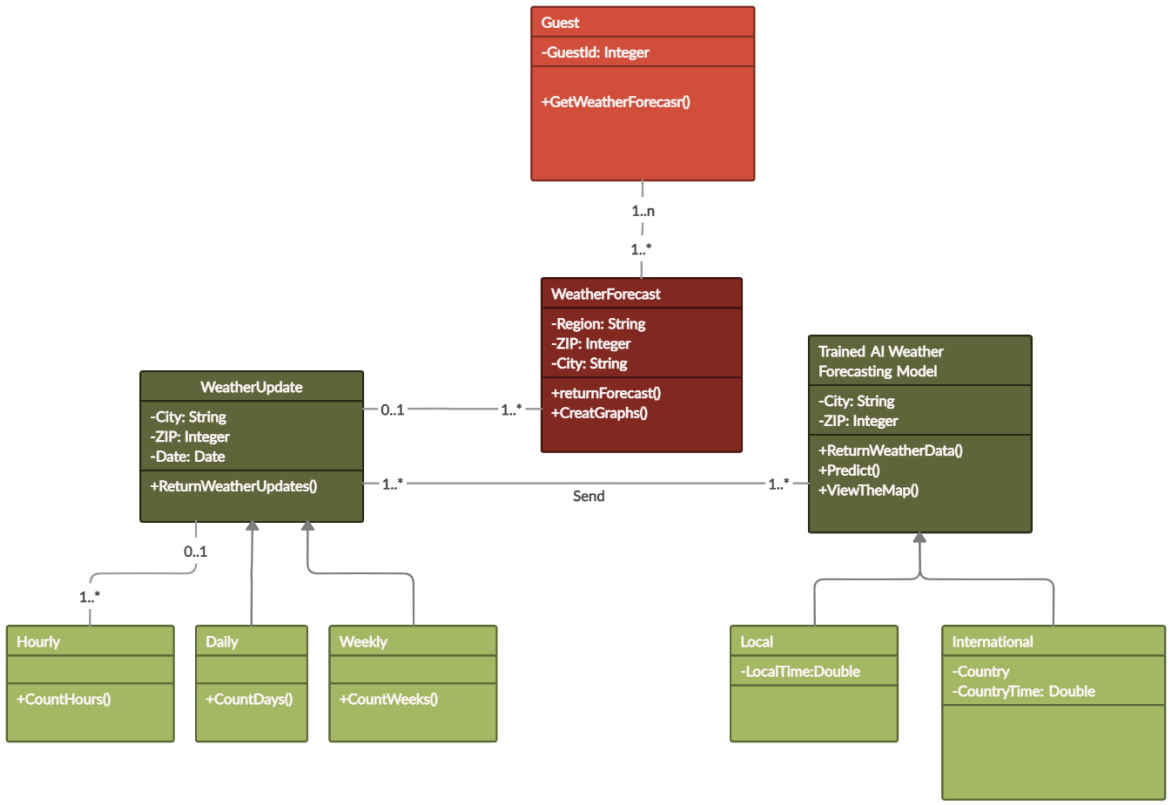
Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

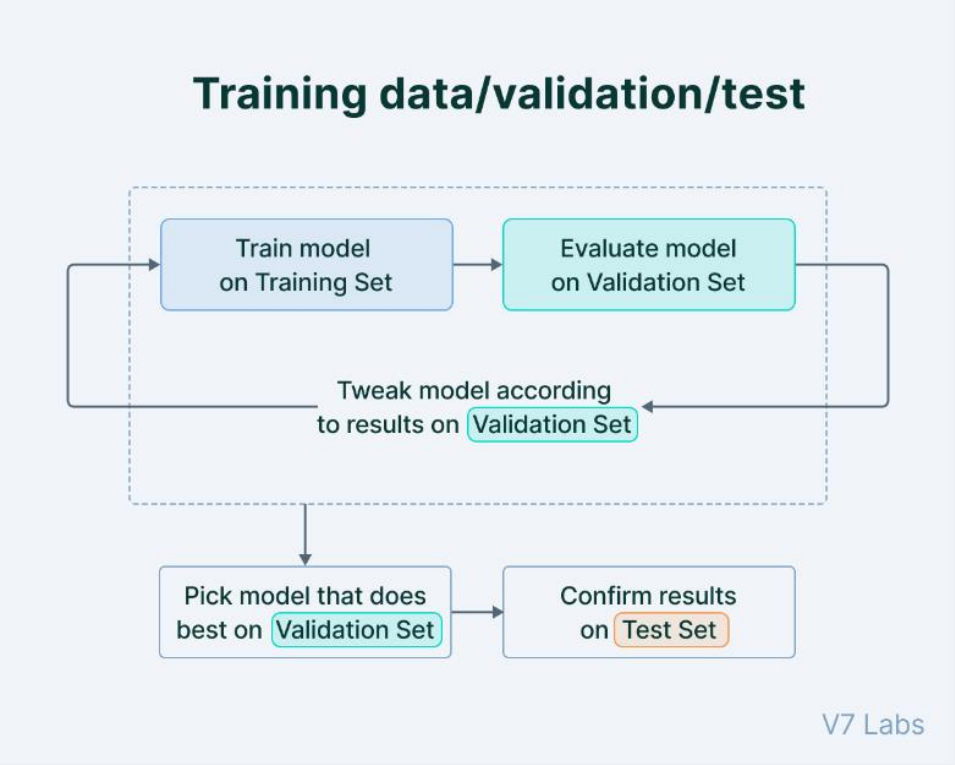
Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

7. System/Software Architecture Diagram

The graph blow described How the webapp functions.



The following graph shows how will I train/test/validate models [3]



To get the train/test/validate data, I will use Sklearn's split method

```
# set aside 20% of train and test data for evaluation
X_train, X_test, y_train, y_test = train_test_split(train, test,
    test_size=0.2, shuffle = True, random_state = 8)

# Use the same function above for the validation set
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train,
    test_size=0.25, random_state= 8) # 0.25 x 0.8 = 0.2
```

First split the data into train and test sets, then split the train set again to get validate set. In this way you will get 60% data for train set, and 20% each for test and validate set.

8. Innovation

The innovation of this project is to improve the speed and accuracy of the current weather forecasting techniques using machine learning. The current weather forecasting apps on the market still use the combination of computer calculation and human expert analysis to predict the weather, but this kind of forecast is not always accurate, and takes hours to calculate. However, with AI forecasting, you can get the result in seconds at a very high accuracy. Let AI model to solve human's problem is the project's main innovation.

Another innovation of AI weather forecasting comes from its ability to make more accurately short-term and long-term predictions. [4] Traditional methods use complex equations and often forecast for only between six hours and two weeks' time, and hardly to forecast for a longer range, but AI can predict if it will rain in two hours' time or after two-six weeks.

Moreover, normal weather forecasting apps don't provide nature disasters forecasting. However, AI weather forecasting system could help identify potential extreme weather 2–6 weeks into the future [5]. Accurate predictions of extreme weather with a longer lead time give communities and critical sectors such as public health, water management, energy, and agriculture more time to prepare for and mitigate potential disasters.

9. Complexity

The project is complex because it requires research skills, machine learning knowledges and software design skills. Diploma students never learned any courses about AI machine learning, only the machine learning option students learned a bit, they won't know how to train, test and validate models. This project also requires a lot of research, for example: research the algorithm and study; find the datasets and cleaning; and look for the solutions when face difficulties during the development. The lack of research skills would cause difficulties for diplomas to start the project. Moreover, diploma students don't have enough knowledge about software design architecture and project management. The projects they have done usually have clear instructions and requirements, but for this project you have to define your own requirements. It would be too hard for them to create such a big project from the scratch. Therefore, only those with a bachelor or higher education can complete this project.

10. Technical Challenges

1. Research for the suitable machine learning algorithm to start with is a challenge: There are many algorithms that can study patterns from images, but their focus is different. After I find a suitable algorithm, I have to study it first, so I can develop upon this algorithm to meet my needs.
2. Clean up the datasets and add tags would cost a lot of time: The datasets found on the internet may not labeled correctly and often has data partially missing. Clean up the data and label them is an important job in model training and therefore will cost a lot time.
3. Train the model and improve its speed and accuracy: The model needs to be trained again and again to be useable. If a model functions out of my expectation, I have to know how to fix it.
4. Design the software and define the milestones are also the challenges. How to manage the workflow and time is difficult when doing large projects. Designing the software architecture is also very complex.

11. Development Schedule and Milestones

The milestones are broken into 5 sprints and each sprint contains two weeks. Each sprint contains a 35 to 110 hours work period, with a total of 365 hours. If the estimated hours are not correct, I will update the real time used in the final report.

Timeline	*Milestones*	*Breakdown*	*Total*
Jan 4th to Jan 18th	**Sprint 1 Development environment configuration** ● **Research for all the required materials** ● **Configure Python environment** ● **Install any required packages/ modules** ● **Play around with installed modules and learn the basic usage**	**Planning: 15hrs** **Coding: 5hrs** **Installation: 5hrs** **Testing: 5hrs** **Documentation: 5hrs**	**35hrs**
Jan 18th to Feb 1st	**Sprint 2 Cleaning and dividing Data into sets** ● **Download Datasets** ● **Cleaning Data** ● **Divide Data into Training, Testing and Validation Sets.** ● **Write a python program to load data and display in graphs**	**Planning: 10hrs** **Cleaning: 10hrs** **Coding: 15hrs** **Testing: 10hrs** **Documentation: 5hrs**	**50hrs**
Feb 15th to Mar 1th	**Sprint 3 Training AI models** ● **Design the models** ● **Implement the R.F. algorithm** ● **Train the model** ● **Adjust the model to increase performance** ● **Able to predict weather**	**Planning: 10hrs** **Designing: 20hrs** **Coding: 40hrs** **Training: 15hrs** **Testing: 15hrs** **Documentation: 10hrs**	**110hrs**
Mar 1th to Mar 15th	**Sprint 4 Testing and Validating models** ● **Adjust model to enhance accuracy** ● **Input testing data to check performance** ● **Input partial missing data to check performance** ● **Input a new set of data for validation**	**Planning: 10hrs** **Coding:15hrs** **Testing:30hrs** **Documentation:10hrs**	**65hrs**
Mar 15th to Mar 29th	**Sprint 5 Implementing the AI models to the webapp and wrapping up**	**Planning: 10hrs**	**75hrs**

	• **Write the basic webapp framework** • **Implement the AI model in the app** • **Make sure the app display correctly on the web and phones.** • **Make sure the webapp run offline** • **Debugging**	**Design:10hrs** **Coding: 30hrs** **Testing: 15hrs** **Documentation: 10hrs**	
Mar 29th to Apr 12th	**Release** **Final Testing** **Final Report**	**Testing: 15h** **Documentation: 15hrs**	**30hrs**
Estimated Total Hours			**365hrs**

12. Deliverables

The deliverables for this project are as follows:

- Weather Forecasting Application: Include a web application with source code.
- Machine learning models: Fully trained models that forecast temperature and humidity.
- Datasets: Training Set, Testing Set and Validation Set.
- Final Report: Final report for the practicum course.
- Demo Videos: Show different test cases and results.

13. Conclusion and Expertise Development

In conclusion, this major project has real value for our daily life, and it is beneficial to enhance my programing skills. The practicum project will help enhance my software development skills such as software architecture design, algorithm design, Python programming, and AI programming. To design the software, I need to have clear understanding on software architecture. During Btech studies, I had a course about software Engineering, this project gives a great opportunity to practice and improve the software design skills I obtained from Btech studies. My Python programming skill will be improved because the application framework will be written in python, and the AI model will also be written in Python. The project will also improve my AI and algorithm programming because I need to implement

and adjust the proper AI algorithm for my model training. It is helpful to my future courses because I have an AI course next term. The project will also help increase my researching skills as I will have to research for the AI programming knowledge and google, stack overflow for answers in case I have troubles during development. In addition, there might be unforeseen challenges discovered during the development. My skills will be improved as I confront the technical challenges of the project and works to overcome them.

14. References

1. Tyler Herrington. (2019, September 23). Why is the weather so hard to predict?

   Retrieved November 2, 2022, from

   https://letstalkscience.ca/educational-resources/stem-in-context/why-weather-so-hard-predict

2. Javatpoint. (n.d.). Random Forest Algorithm

   Retrieved November 2, 2022, from

   https://www.javatpoint.com/machine-learning-random-forest-algorithm

3. Pragati Baheti. (2022, October 21). Train Test Validation Split: How To & Best Practices [2022]

   Retrieved November 2, 2022, from https://www.v7labs.com/blog/train-validation-test-set

4. BBC. (2021, September 30). AI can predict if it will rain in two hours' time

   Retrieved November 2, 2022, from https://www.bbc.com/news/technology-58748934

5. Michelle Horton. (2022, January 13). Global AI Weather Forecaster Makes Predictions in Seconds

   Retrieved November 2, 2022, from

   https://developer.nvidia.com/blog/global-ai-weather-forecaster-makes-predictions-in-seconds/

6. Hannah Hickey. (2020, December 15). A.I. model shows promise to generate faster, more accurate weather forecasts

   Retrieved November 2, 2022, from

https://www.washington.edu/news/2020/12/15/a-i-model-shows-promise-to-generate-faster-more-accurate-weather-forecasts/

7. Piero Paialunga. (2021, Apr 19) Weather forecasting with Machine Learning, using Python

Retrieved November 2, 2022, from

https://towardsdatascience.com/weather-forecasting-with-machine-learning-using-python-55e90c346647

8. Marwah Sattar Hanoon, Ali Najah Ahmed, Nur'atiah Zaini, Arif Razzaq, Pavitra Kumar, Mohsen Sherif, Ahmed Sefelnasr & Ahmed El-Shafie (2021, September 23). Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu state in Malaysia

Retrieved November 2, 2022, from https://www.nature.com/articles/s41598-021-96872-w