

School of Public Administration  
Bachelor of Science in Computing

**COMP491 Final Year Project  
Progress Report**Academic Year 2018/19

|  |  |
| --- | --- |
| Macao Bus Travel Time Prediction Using Neural Network | |
|  |  |
| Project number: | 24 |
| Student ID: | P1507881 |
| Student Name: | Chris Zhou |
|  |  |
| Supervisor: | Benjamin |
| Assessor: | June Liu |
|  |  |
| Submission Date: | 2018.11.15 |

Declaration of Originality

I, Chris Zhou, declare that this report and the work reported herein was composed by and originated entirely from me. This report has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given in the bibliography.



11.11.2018

Abstract

In Macao, thousands of commuters travel to work or school by taking buses every day. It is known that a city like Macao is often plagued by traffic jam and huge demand in public transit, which result in bus delay and long bus waiting time. This is particularly hectic for people travelling in peak hours. In order to solve this kind of problem, this project is to develop an application allowing the users to predict the travel time required taking into account various conditions, such as the time of the day, weekend/weekday, the weather conditions (good weather, bad weather), the route and so on. After the user enter these input parameters, the application will give an output number which is the prediction travelling time. The outcome may help the users to weight different travel options and prepare beforehand.

Acknowledge

During the 4-years college life in Macao Polytechnic Institute, all of my teachers, friends and classmates help me a lot. I am very appreciated that my friends always support and comfort me when I feel stressful and negative. I feel grateful for my teachers teaching me the knowledge and offering some suggestions for my career, I would never forget all your great kind-hearted and the knowledge I learnt. I am thankful for my classmates who always stay with me during these years and help each other on study. I'm not an expressive man, but all of you will be remembered for the rest of my life. In this final year project, I have got huge amount of help and advice from my supervisor Benjamin Ng, thank you for your suggestions and support, I cannot keep this project ongoing without you.

Table of Contents

[1 Introduction 8](#_Toc2800191)

[1.1 Objectives 9](#_Toc2800192)

[1.2 Risk Assessment 9](#_Toc2800193)

[1.3 Summary 11](#_Toc2800194)

[2 Background and Related Work 12](#_Toc2800195)

[2.1 Population and public transport in Macao 12](#_Toc2800196)

[2.2 Big Data 14](#_Toc2800197)

[2.3 Neural Networks 15](#_Toc2800198)

[2.4 The Main Software Tools Used 18](#_Toc2800199)

[2.4.1 Pybrain 18](#_Toc2800200)

[2.4.2 Bootstrap 19](#_Toc2800201)

[2.4.3 Flask 19](#_Toc2800202)

[2.5 Related Work 20](#_Toc2800203)

[3 Completed Work 22](#_Toc2800204)

[3.1 System Structure 22](#_Toc2800205)

[3.2 Dataset 23](#_Toc2800206)

[3.3 Neural Network 23](#_Toc2800207)

[4 On-going and Future Work 24](#_Toc2800208)

[4.1 Server 24](#_Toc2800209)

[4.2 Interface 24](#_Toc2800210)

[5 Conclusion 25](#_Toc2800211)

[References 26](#_Toc2800212)

Table of Figures

[Figure 1: Probability impact matrix 9](#_Toc530079378)

[Figure 2 Total population in Macao 11](#_Toc530079379)

[Figure 3 The natural population growth rate 12](#_Toc530079380)

[Figure 4 Population intensity 12](#_Toc530079381)

[Figure 5 Total registered motor vehicle 13](#_Toc530079382)

[Figure 6 The number of cars per thousand people in Macao 13](#_Toc530079383)

[Figure 7 a single neuron 15](#_Toc530079384)

[Figure 8 activation functions 16](#_Toc530079385)

[Figure 9 feedforward neural network 16](#_Toc530079386)

[Figure 10 Pybrain's architecture 18](#_Toc530079387)

[Figure 11 4 Components of the project 21](#_Toc530079388)

[Figure 12 activity diagram 22](#_Toc530079389)

[Figure 13 part of dataset 23](#_Toc530079390)

[Figure 14 make the dataset suitable for neural network 24](#_Toc530079391)

[Figure 15 build the network and train it by Backpropagation algorithm 25](#_Toc530079392)

[Figure 16 initialization and test it 25](#_Toc530079393)

[Figure 17 error rate and result 26](#_Toc530079394)

[Figure 18 Gantt Chart 27](#_Toc530079395)

List of Tables

[Table 1: Table of prioritized risk 8](#_Toc530079396)

# Introduction

As is known to all, Macao is the city which has the highest population density in China (Wiki, 2018) [1], and it is famous for its thriving tourism and gambling industry. Therefore, Macao attracts millions of people. For example, a great number of travellers coming to Macao for fun, thousands of gamblers coming here for make big money, and a good deal of people coming for a better job or higher salary. However, Macao is not a big city like Beijing. It has a limited range of land; narrow loads usually have one or two lanes. Thus, the arrival of so many people makes the city comparatively crowded, which causes a big trouble for the public transport and there is more and more traffic jam. Also, there are various factors which affect the public transport, such as weather and the time of the day. Hence, due to the unstable bus travelling time, people in Macao have a huge demand for something which is able to predict the bus travelling time regarding to some factors which influence transport.

To solve this kind of problem, there are various mobile application helping users to check the public transport information. Some of them only offer information of bus route, bus stop, real-time bus location and so on. However, it is not able to make any prediction. Macao Bus Traveling System [2] is a typical example. Public Transport Victoria (PTV) [3] provide users a plan maker. After choosing a start location and destination, users are able to plan the journey in very detail, and the travelling time will also show up according to the timetable. But the estimated travelling time is absolutely based on a fix timetable. 8684.cn [4] is a public transport searching system, including the travelling information of bus, train, and plane. It supports most cities in China. But it does not support public transport of Macao. Maybe the reason is that it cannot connect the database in Macao.

To cover the aforementioned shortages, this project aims to provide a comparatively perfect application for Macao people. It is to develop an application allowing the users to predict the travel time required taking into account various conditions, such as the time of the day, weekend/weekday, the weather conditions (rain, heavy rain, typhoon, etc.), the route and so on. The outcome may help the users to weight different travel options and prepare beforehand.

This project’s expected result is mainly a proof-of-concept prototype, and as such, the estimation needs not be very accurate and may be subject to modification in the future once more data is available to train the neural network [5].

## Objectives

There are two main objectives in the project. First, build a web-based application with an interface allowing users to enter the travel information and the traffic parameters (such as time, weather conditions and so on). Second, build a neural network which accepts the traffic parameters and produces an estimation of travel time including the bus waiting time. **The major requirements for this project are listed below:**

* Building a user-friendly interface
* Collecting the sample data from the application “Bus Traveling System”, then training the neural network by these sample data
* Users are able to choose a bus route, starting location and destination and the traffic parameters
* Submitting the input from user to the neural network
* Output the prediction from neural network and show it in the interface

## Risk Assessment

Table 1: Table of prioritized risk

|  |  |
| --- | --- |
| Priority | Risk Identifier and Description |
| 1 | Risk 1: The project files are disappeared and there is no backup file |
| 2 | Risk 2: The student or the supervisor has sick |
| 3 | Risk 3: The user’s web browser is not compatible with the application |
| 4 | Risk 4: The result prediction is not precise |

Note: Priority 1 is the highest risk

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Probability** | High |  | Risk 4 |  |
| Medium |  | Risk 3 |  |
| Low |  |  | Risk 1  Risk 2 |
|  | Low | Medium | High |
| **Impact** | | |

Figure 1: Probability impact matrix

Risk 1: The project files are disappeared and there is no backup file. The application may not work at all. It takes a long time to redo it.

Solution: Making a backup at regular intervals.

Risk 2: The project supervisor or the student doing the project may get sick one day. It’s hard to do the work when sick.

Solution: Have a few days rest to recover. It’s necessary to keep safe and keep healthy, and do the project as soon as possible when there is no sick.

Risk 3: There are so many browsers in both PC and mobile phone, the web-based application may not be compatible for all the browsers. For example, the application can work perfectly on Chrome but some content may miss or the layout is changed in Firefox.

Solution: Try three most popular browsers, see if all of them can work perfectly. If not, fix the bug as soon as possible.

Risk 4: The result prediction is not precise. But the prediction usually won’t go far from the actual bus travelling time.

Solution: Have more training to the neural network in order to make it more precise.

## Summary

This progress report is divided into 5 chapters. Chapter 1 defines the project with clear objectives. It contains the project motivation and the problem of another existing works. Besides, the risk is also discussed in this chapter. Chapter 2 states the background of this project as well as the detail related works. Chapter 3 presents the work which is already done in the first semester. Chapter 4 states the future work in next semester. And there is a conclusion in the Chapter 5.

# Background and Related Work

With the development of internet and smartphone, the use of bus travelling time prediction application is getting popular. In this chapter, the basic facts about population and public transport in Macao will be illustrated. The main technologies used in this final year project and other similar programs will be introduced below.

## Population and public transport in Macao

All the figures below are taken from the statistical database of DSEC [6] (Statistics and Census Service) in Macao.

As shown in Figure 2, the total population in Macao is steady increased by approximately 113 thousand in recent years, which is a huge amount.

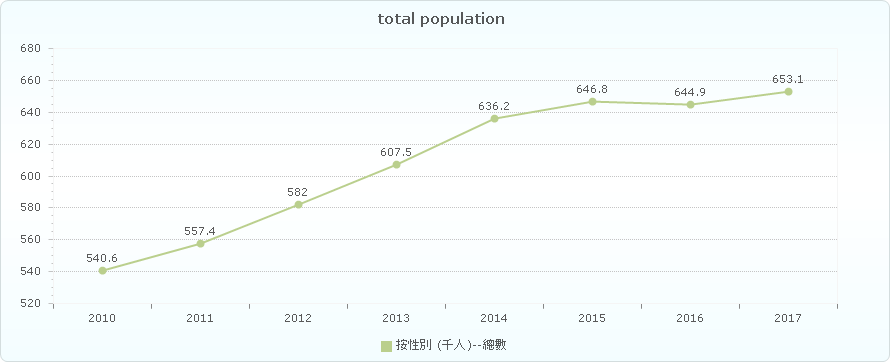


Figure 2 Total population in Macao

As seen in Figure 3, the natural population growth rate in Macao is slightly decreased recently. But even the growth rate is declined, the population of Macao is still increasing rapidly.

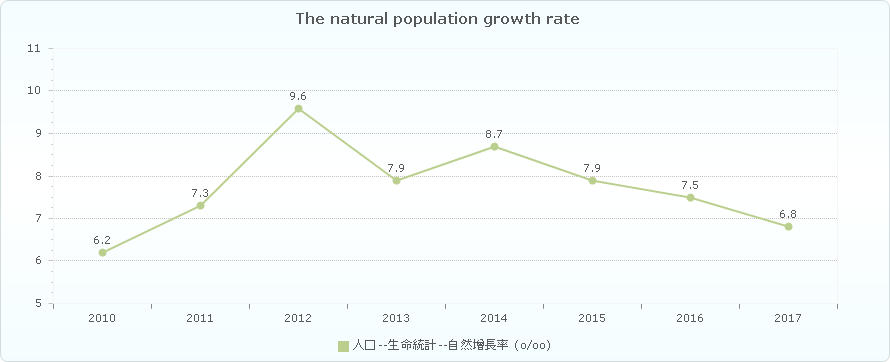


Figure 3 The natural population growth rate

Figure 4 illustrates the population intensity in Macao is always rising from 2010 to 2016. It means that the small city is getting even more crowded.

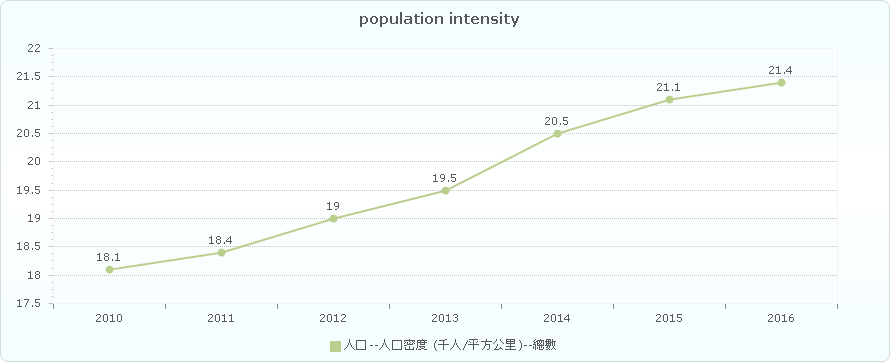


Figure 4 Population intensity

As shown in Figure 5, there are more and more motor vehicle in Macao over six years.

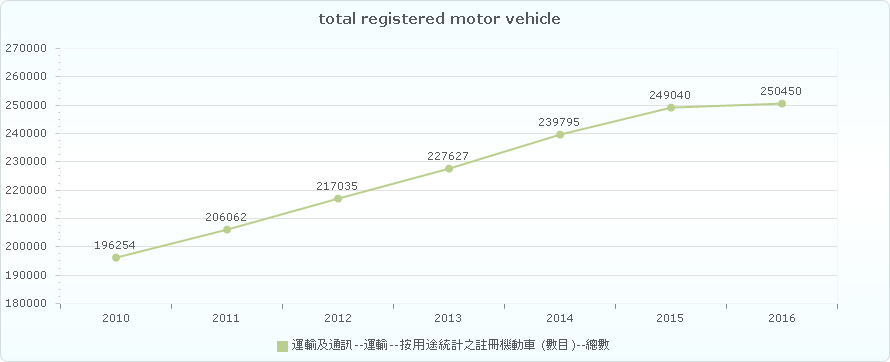


Figure 5 Total registered motor vehicle

As seen in Figure 6, the number of cars per thousand people in Macao increased steadily, which making the traffic more crowded.

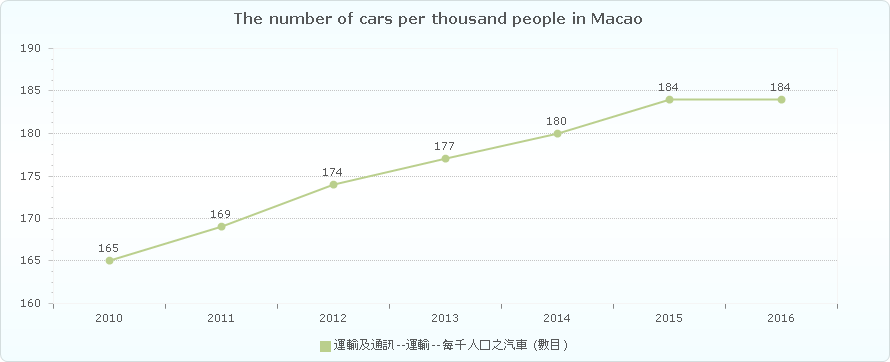


Figure 6 The number of cars per thousand people in Macao

## Big Data

Big data [7] is a word which is used to refer to dataset that are too complex or large for traditional data-processing software to sufficiently deal with. But it’s not the volume of data which matters. It is how people do with the data that is important. Big data can be analysed for discernments which may lead to better decisions and strategic business movement. Big data contains greater variety arriving in increasing volumes and with ever-higher velocity. This is known as the three Vs [8].

* Velocity. Velocity means the speed rate at when data is received or transmitting. Usually, the highest velocity of data streams directly into memory versus being written to disk. Some smart products which are able to connect internet, handle in real-time or nearly real-time. Besides it will need real-time evaluation and action.
* Variety. Variety means the various kinds of data which are available. Traditional data types were just orgranised and fit in database. With the increasing development of big data, data may have new unstructured data categories. Extra pre-processing is needed for unstructured data types need to derive meaning.
* Volume. The amount of data is important. This may be data containing unknown value, such as click number of streams of a website, Facebook data, or just a mobile app. The data might be billions of gigabytes of data, or millions of terabytes.

## Neural Networks

An Artificial Neural Network (ANN) [9] is an information processing model which is inspired by the biologic brain, for example how people’s brain manages transactions and information. The most important part of this model is the innovative architecture of the information processing system. It consists of a huge amount of interactive processing elements (which is called “neurones") working together to deal with some certain problems. ANN is similar with human, it can learn things by given sample. An ANN is set for a certain program, such as data classification or pattern recognition, by the process of learning. Learning in biology need to make adjustments to the connections which are between the nodes. This theory is also correct in ANNs.

The basic unit of computation in a neural network is the neuron [10], or called a node. The input is from an outside source or some other nodes, then calculates a result. Each input has a related weight (which is the letter “w” in Figure 7). The neuron has a function as “f “, which is defined below in the Figure 7.

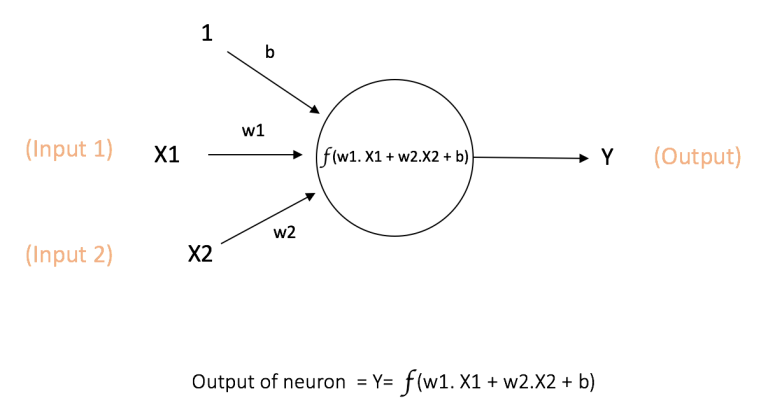


Figure 7 A node [10]

The above figure shows the inputs X1 and X2, which has its own weights w1 and w2 together with these inputs. Besides, there is additional input 1 and its weight b which is called the “Bias”.

The result Y is calculated by the function “f” which can be seen in the Figure 7. The function “f” is non-linear which is the “Activation Function”. The activation function aims at non-linearity. It matters because the most data in reality is non-linear and neurons needs to learn these non-linear representations.Every activation function takes a single number and performs a fixed mathematical operation on it. Here are some popular activation functions which are useful:

* **Sigmoid**: σ(x) = 1 / (1 + exp(−x))
* **tanh**: tanh(x) = 2σ(2x) − 1
* **ReLU**: f(x) = max(0, x)

The Figures 8 [10] below shows the activation functions above.

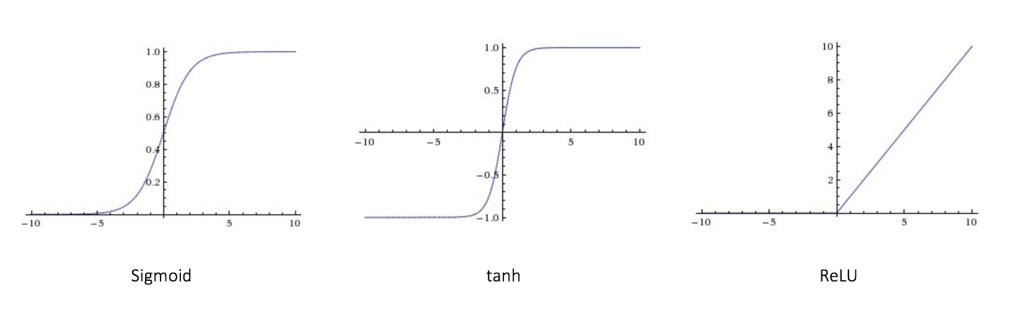


Figure 8 activation functions

As shown in figure 9, the feedforward neural network (FNN) was the easiest kind of artificial neural network (ANN). It has many nodes in each layer. The nodes in neighbour layers have connections or edges between layers. Every connection has its own weights.

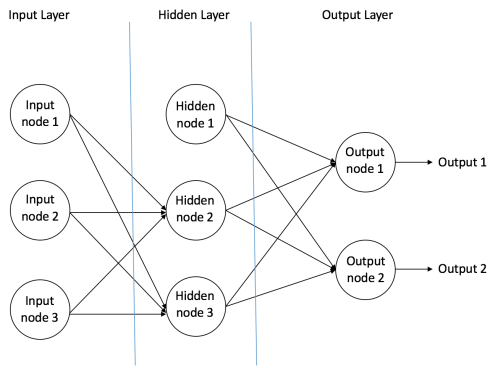


Figure 9 feedforward neural network [10]

In a feedforward network, there is only one direction which is forward, starting from the input neuron, then going through the hidden nodes and then finally to the output nodes. There are no circles or loops in feedforward network.

After building up a neural network, it needs to be trained. Backpropagation algorithm is one of the best methods to train the network. At the beginning, all the weights are assigned randomly. Taking all input in the training dataset, the ANN is available and its output can be known. This output will be compared with the ideal output which we are already aware of, and then the error is “propagated” back to the previous layer. The error is marked, and then the weights will be “adjusted” accordingly. The process is done again and again until the error of the output is below a predetermined threshold. After this Backpropagation algorithm process finishes, a “trained” neural network is done, which we believe it is good enough to work with some “new” inputs.

## The Main Software Tools Used

### Pybrain

PyBrain [11] is a modular Machine Learning Library for Python. It aims to offer flexible, easy-to-use and still stronger algorithms for Machine Learning Tasks and a variety of predefined environments to test and compare your algorithms. PyBrain is short for Python-Based Reinforcement Learning, Artificial Intelligence and Neural Network Library.

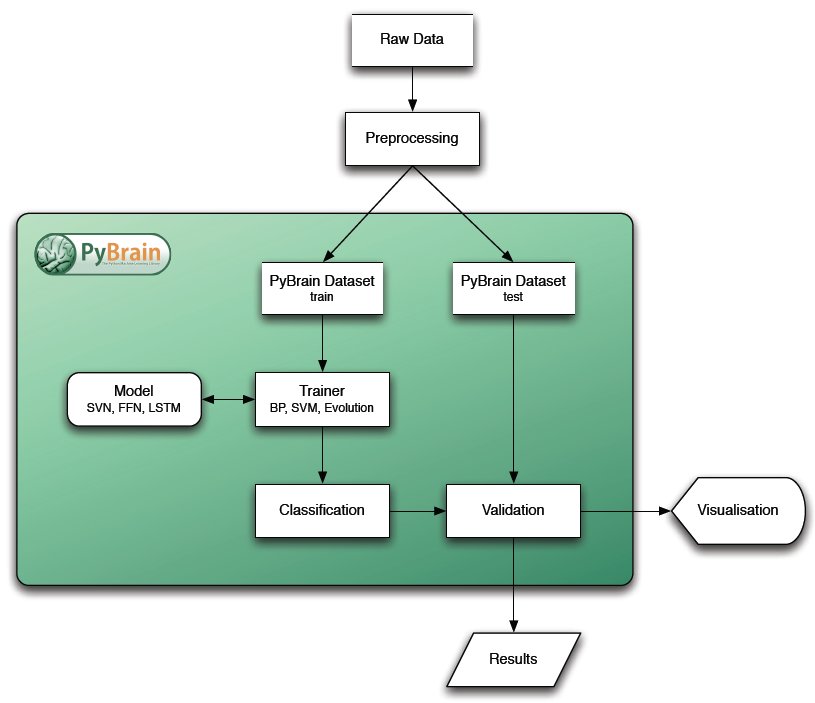


Figure 10 Pybrain's architecture

### Bootstrap

Bootstrap is a free and open-source front-end Web framework. It contains HTML and CSS-based design templates for typography, forms, buttons, navigation and other interface components, as well as optional JavaScript extensions. Unlike many earlier web frameworks, it concerns itself with front-end development only.

### Flask

Flask is a micro web framework written in Python. It is classified as a micro framework because it does not require particular tools or libraries (except for some basics standard libraries such as bottom.py. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common frameworks related tools. Extensions are updated far more regularly than the core Flask program. [5] Flask is commonly used with MongoDB, which gives it more control over databases and history.

## Related Work

With the development of internet, web-based application can be accessible for both PC and mobile phone. Thus, internet programming is an important skill for this project. In order to get familiar with the neural network, the knowledge of neural network is a must. Also, programming languages such as python is needed to complete the neural network.

2.5.1 Bus Traveling System

The application “Bus Traveling System” [2] will be used in this project. It’s officially released by DSAT [13], in order to help people in Macao knowing the information about each bus. It provides information of bus route, bus stop, real-time bus location and so on. It contains huge amount of data and it is helpful for citizen in Macao. However, it cannot predict the bus travelling time regarding to the real traffic condition, and there is no input to adjust the result prediction. Therefore, the project’s expected result is able to make the bus travelling time prediction much more precise according to the real traffic condition, allowing users to input the parameters (weekday/weekend, weather and so on), and give the prediction time dynamically according to these input parameters.

2.5.2 Public Transport Victoria (PTV)

Public Transport Victoria (PTV) [3] is a statutory authority that manages train, tram and bus services in Victoria, Australia. It provides a single contact point for you to gain information on public transport services, fares, tickets and initiatives. It contains the whole Victoria’s public traffic information. After choosing a start location and destination, users are able to plan the journey in very detail, and the travelling time will also show up according to the timetable. However, the travelling time is absolutely based on a fix timetable. It cannot change dynamically regarding to the real traffic condition. Sometimes people in Victoria complain about the traffic because the buses always delay while they don’t even know when the bus is coming.

2.5.3 8684.cn

8684.cn [4] is a public transport searching system, including the travelling information of bus, train, and plane. It supports most cities in China. After users set start location and destination, the system will show the plan in detail. However, it doesn’t show the travelling time for users, only the plan and direction, Due to the difficulty of connecting to all cities’ traffic database in China.

# Completed Work

This chapter is going to describe the completed work of the project. First, the system architecture of the project will be illustrated, besides the ideal user activity diagram. Then the dataset and a workable neural network will be introduced.

## System Structure

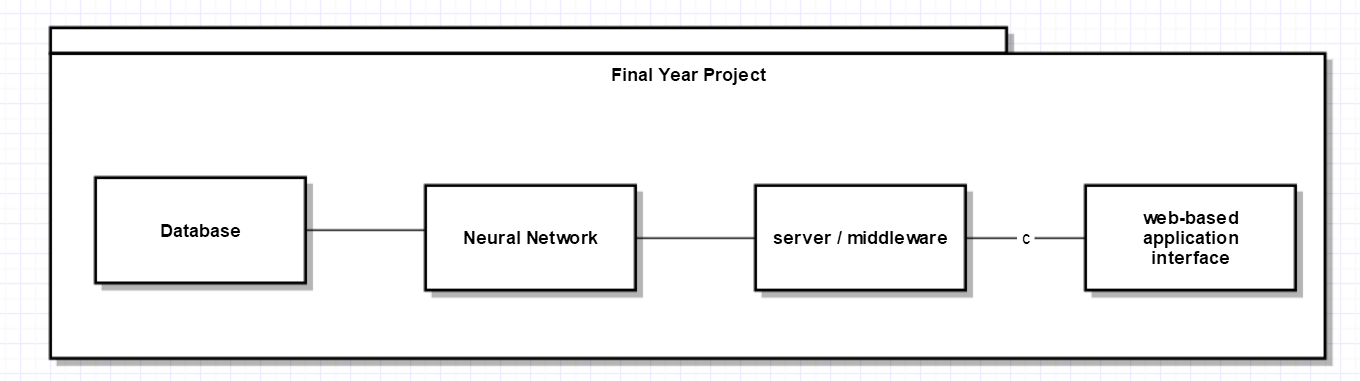
Basically, there are four components in this project. As seen in the figure 11, there are Database, Neural Network, Server and the web-based application interface. Data is collected from the mobile application “Macao Bus Travelling System”. Then it is normalized into a csv file. The neural network is made by python with PyBrain library, and it is trained by the previous dataset. The server is built for connecting the web application with neural network. Finally, the interface is going to be made to make it easier for users.

Figure 11 4 Components of the project

Figure 12 is the activity diagram, showing the basic usage of this project. Firstly, collecting data and build the neural network. Then training the neural network by this dataset. When the neural network is available, users are able to use the application, inputting the parameters and get the result prediction with the help of neural network.

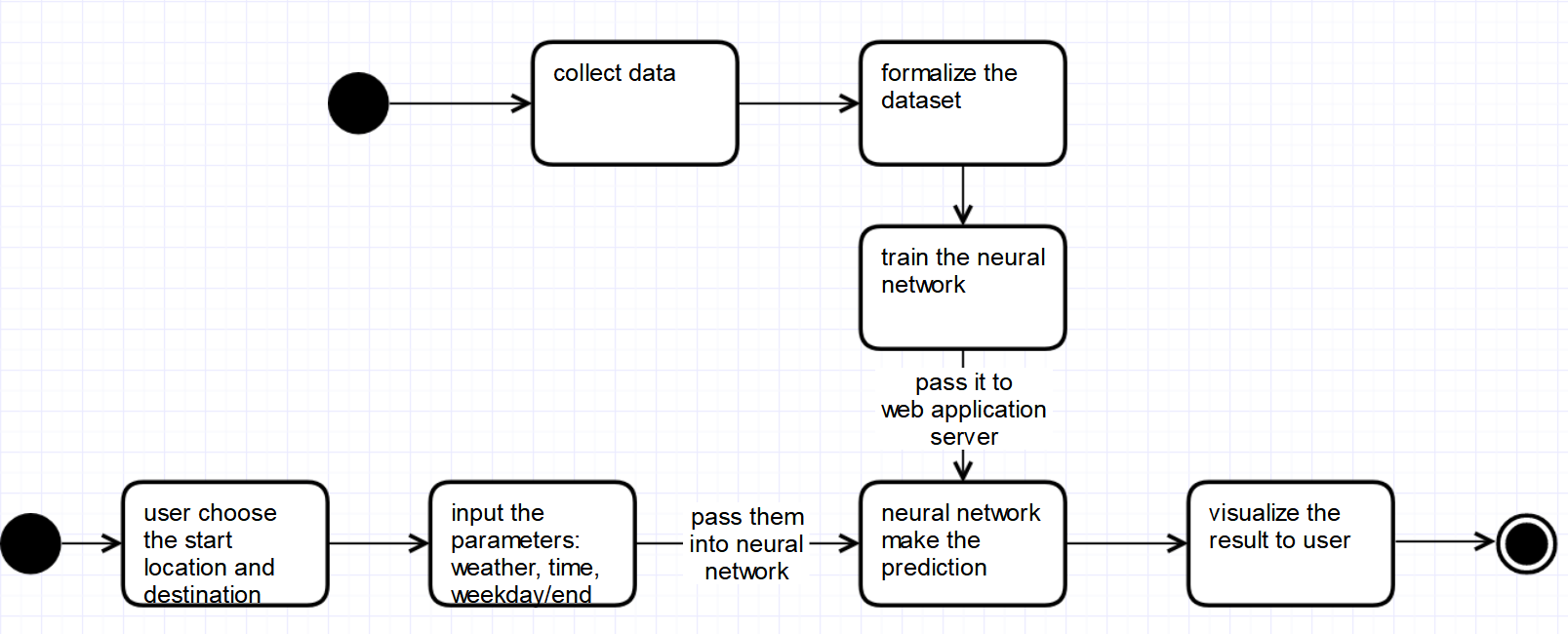


Figure 12 activity diagram

## Dataset

As shown in Figure 13, these is a part of my collected dataset. It costs pretty long time to collect data manually so the process of collecting data hasn’t finished, and It might continue to next semester because if the dataset is larger, the neural network may become more precise.

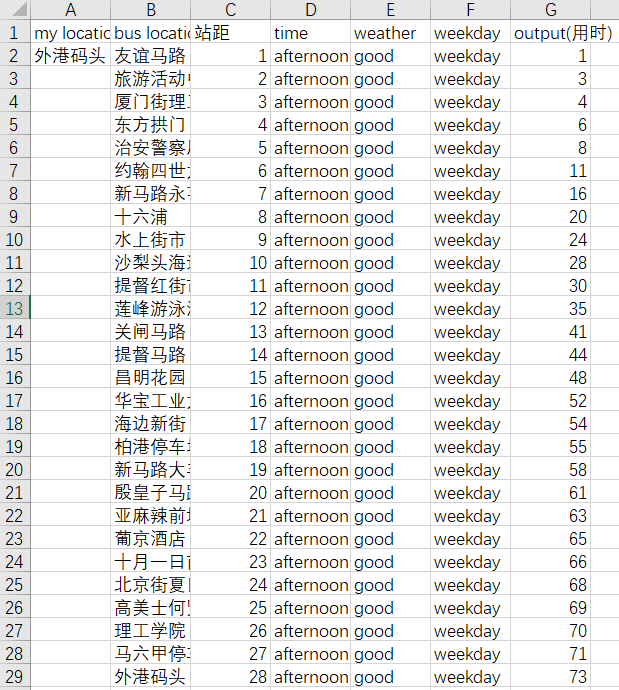


Figure 13 part of dataset

## Neural Network

# On-going and Future Work

According to Figure 11, the database part is partially finished as well as the most difficult part--neural network. In the future, the other two components, server and interface need to be done.

Figure 18 is the Gantt Chart in Nov.15.2018.

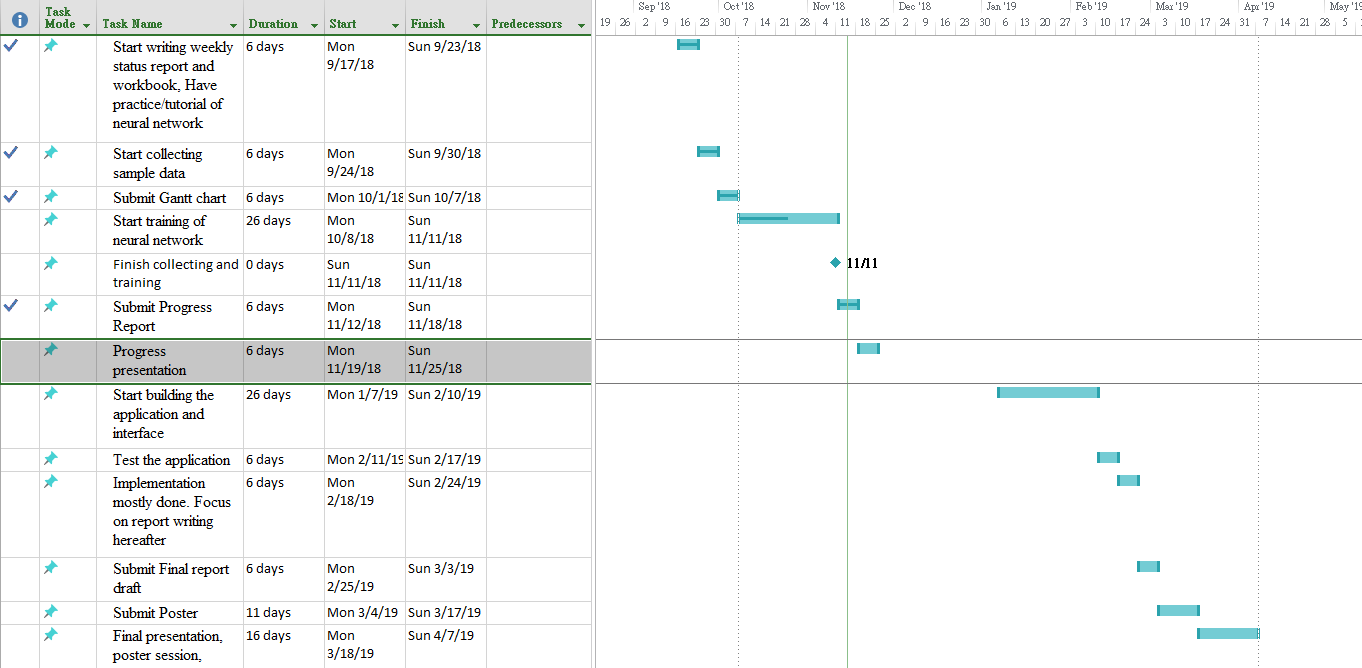


Figure 18 Gantt Chart

## Server

In order to deploy the server quickly and easily, XAMPP [15] is selected, a package contains Apache [16] deployment, PHP [17] and MySql [18]. This is quite a simply way to setup a server and development environment.

## Interface

As for the interface of the web-based application, PHP is selected because the previous package container XAMPP [15] has already setup the environment. Also, PHP is one of the most popular language for building a website.

# Conclusion

The reason that I select the topic is the neural network, which is the new technology I never seen and never been taught before, and it is a hot topic recently indeed. In the first semester, I am aware of the importance of self-learning and self-motivation. Unfortunately, I suffer the injury of ankle twisting in October, which resulted in the inconvenient of movement. Therefore, I missed some of meeting with my supervisor and delayed the project somehow according to the project plan. However, I made a workable neural network and I think this is the most difficult part in this project. Anyway, I need to push harder to catch up the schedule in the future.

References

[1] Wikipedia. List of cities by population density. <https://en.wikipedia.org/wiki/List_of_cities_by_population_density>. [Nov. 11, 2018].

[2] DSAT. Main page of Transport Bureau. <http://www.dsat.gov.mo/dsat/index.aspx>. [Sep. 11, 2018].

[3] Victoria Government. Main page of Public Transport Victoria. <https://www.ptv.vic.gov.au>. [Sep. 11, 2018].

[4] Guangzhou Tianqu. Main page of 8684. <http://www.8684.cn>. [Sep. 11, 2018].

[5] Wikipedia. Artificial neural network. https://en.wikipedia.org/wiki/Artificial\_neural\_network. [Sep. 11, 2018].

[6] DSEC. Statistics and Census Service. <https://www.dsec.gov.mo/home_zhmo.aspx>. [Nov. 11, 2018].

[7] SAS. What is big data and why it matters. https://www.sas.com/en\_us/insights/big-data/what-is-big-data.html. [Nov. 11, 2018].

[8] Oracle. The Three Vs of Big Data. <https://www.oracle.com/big-data/guide/what-is-big-data.html>. [Nov. 11, 2018].

[9] Christos Stergiou and Dimitrios Siganos. Introduction to neural networks. https://www.doc.ic.ac.uk/~nd/surprise\_96/journal/vol4/cs11/report.html#What is a Neural Network. [Nov. 11, 2018].

[10] Ujjwalkarn. A Quick Introduction to Neural Networks. <https://ujjwalkarn.me/2016/08/09/quick-intro-neural-networks/>. [Aug.9.2016]

[11] Tom Schaul, Justin Bayer, Daan Wierstra, Sun Yi, Martin Felder, Frank Sehnke, Thomas Rückstieß, Jürgen Schmidhuber. PyBrain. <http://pybrain.org/pages/home>. [Nov. 11, 2018].

[12] INRIA. scikit-learn, Machine Learning in Python. <https://scikit-learn.org/stable/>. [Nov. 11, 2018].

[13] Dreijer, Janto. scikit-learn. <https://scikits.appspot.com/scikit-learn>. [Nov. 11, 2018].

[14] scikit-learn developers. History of scikit-learn. <https://scikit-learn.org/stable/about.html#history>. [Nov. 11, 2018].

[15] XAMPP community. What is XAMPP? <https://www.apachefriends.org/index.html>. [Nov. 11, 2018].

[16] Apache Software Foundation. Apache Home page. <https://www.apache.org/>. [Nov. 11, 2018].

[17] The PHP Group. Home page of PHP. <http://php.net/>. [Nov. 11, 2018].

[18] Oracle Corporation. Home page of MYSQL. <https://www.mysql.com/>. [Nov. 11, 2018].