Data Science and AI

Final Exam Alternative

Revision Tutorial Logbook

A picture containing logo

Description automatically generated

Date | 23rd April 2021

Student | Jack McNally

Date Due | C00228758

Lecturer | Greg Doyle

Table of Contents

[December 2020 3](#_Toc69936298)

[January 2021 3](#_Toc69936299)

[February 2021 4](#_Toc69936300)

[March 2021 4](#_Toc69936301)

[April 2021 5](#_Toc69936302)

# December 2020

* This is the first log for the data science and AI project.
* The first thing we have to decide is what algorithms we plan on using.
* After some thinking and from the lectures I am going to implement the algorithms as we learn them in the class. From what I have heard in the lectures the first machine learning language we are going to learn about is decision trees, I think I will pick that to begin with.
* The first deliverable is due in January and this is a detailed plan of some sort that basically describes the whole revision tutorial.
* I have been thinking of using Jupyter notebooks lately because I used them in the group project and they we handy to use.

# January 2021

* Just got back from the Christmas break going to pick up on this work now. The detailed plan is due in a week.
* I have noted in the plan that I am going to use a combination of word files to display the theory side and I’m going to use Jupyter notebooks to bring the code to life. I also discussed the 6 points that were on the specification and how I was going to structure each algorithm to them points.
* I feel like this the content in this document will change as I begin implementing the algorithms but I’m happy with the work I’m going to submit to Greg.
* Just submitted the detailed plan didn’t make a lot of changes to the previous log I’ve made here. I have a lot of work for a lot of different modules at the moment and I’m struggling to find time to do this project, I think I’m going to spend some time at the weekend try to start my first algorithm, decision trees.
* Started the actual Jupyter notebook today and began doing some code, to be honest I didn’t get a lot of work done today It was mostly doing a lot of set up and making sure everything is working, I had an error with the install with Jupyter notebooks I had to install anaconda. I also had quite a problem trying to find a dataset for this problem, I ended up having a meeting with Greg and asking him If could use the dataset and he said it was ok once I used a larger on for the next algorithm.

# February 2021

* Finally got a good day where I got to do a lot of work on the project. I nearly have the decision tree algorithm finished; I nearly have the word document done to. I feel like if I plan this properly and other work doesn’t get in the way I should have this completed.
* It’s the end of February now and I have the decision tree algorithm finished I have decided that the next algorithm I am going to implement is the random forest algorithm. I have been caught up a lot recently with a lot of different work for different modules and once again I am struggling to find time to do work on the next algorithm. Hopefully I can try get some work done this weekend and try have the next algorithm done soon.
* So, I got some time to do a bit of research on the random forest algorithm and I have started a small amount of work, not a lot but enough to get the ball rolling.

# March 2021

* The first draft is due at the end of the month and it’s a week into march I have decided to convert all my work onto a single Jupyter notebook as I heard a few people in class discussing this. I think it will make my work easier to mark and a lot all the information will be in the one place, this will probably take quite a while to convert but will be worth it.
* The plan is due in a week and I’m just about finished the random forest I had quite a tough time on this algorithm, but I think I’m at the final stages soon, I’ve had errors with syntax and data pre-processing.
* The first draft is due tomorrow and unfortunately, I have had to request some extra time to submit the detailed plan, I had a bit of personal issues, but I hope to submit it by the end of the weekend.
* I managed to get the report submitted the day after the deadline. I have now finished to algorithms to a high enough level, I have two more algorithms to implement, and they are going to be Knn and Neural networks. We now have 2 weeks off for and I hope to get Knn done by the end of the break and maybe some of neural.

# April 2021

* So I’m one week into the easter break and I’ve all but finished the Knn algorithms, It took me shorter than I expected I think after the work I did in march really helped really get me in the algorithms frame of mind.
* The decision to move everything into one notebook has been a great decision I find that working is going much quicker because I don’t have to keep switching between programs. I don’t really think I’m going to get a lot more working done on this document before the end of the break as I have a lot of work to done on the project and other assignments we gave due.
* I haven’t updated this log book since the first week of the easter break it is now just about the middle of April I’ve just began Implementing the final algorithm I have chosen which is neural networks. Yesterday I spent all day trying to actually implement the algorithm but unfortunately my virtual box didn’t have enough computing power to run the dataset as it is quite large. So after thinking to myself for quite a while I decided to try download python and Jupyter notebook onto my laptop after a lot of messing around and a lot of fails I couldn’t get it working, so I stared looking for other solutions and eventually fell upon the anaconda navigator.
* This program is fantastic, I am able to do everything on this one platform. From installing libraries to running the complex neural networking algorithm.
* So the project is due in around less than 2 weeks and I’m just about finished. I have some small changes to make here and there but overall I am very happy with the work that I have completed up to now. The use of the anaconda navigator has made it much easier for me to access my work as beforehand I had to run a virtual box to open the work. The plan for the next week is to update this logbook as it is quite a mess in terms of structure as well as record the final demo presentation.
* I also need to done some visualisations to make the notebook that bit more appealing to the eye. Other than all that stuff I have to do the notebook will be finished, to be quite honest this has been quick and slow at the same time and when we receive this project back in December I never thought this deadline could come.
* It’s now the Tuesday before the project is due I have spent all week basically doing the spell checking and making sure the content is correct and in order. I spent quite a bit of time trying to do some visualisations and eventually got that working with a program called plotly. The plan for the rest of the week is to upload all of the datasets onto GitHub so they can be accessed from a safe place. Other than that I have to record the video which is 5 minutes and is worth 10% of the marks.
* Hopefully this is one of my final submissions and when I update this next my project is ready to be submitted on Friday morning.
* So it’s the Thursday the day before the project Is due I have just recorded the video but the file is too large after zipping so I may enquire about it tomorrow. All the work is now completed and I am ready to submit it once I find out how to upload the video. This project has been a great learning experience and I have benefited from it a lot.

Signed: Jack McNally Date: 22/04/2021