**Data Science Project 7PAM2002-0509-2023**

**Semester C 2023**

**Logbook (Activities and GitHub submissions)**

**Student Name and ID: 21061898**

**Project Title: Prostate Cancer Detection**

**Supervisor: Vandana Das**

**Student GitHub URL:** [**https://github.com/Bsyamsundar/Prostatexproject.git**](https://github.com/Bsyamsundar/Prostatexproject.git)

**Number of versions of the code submitted on GitHub:**

**User documentation has been submitted on GitHub: YES / NO**

**Student GitHub URL has been shared with markers: YES / NO**

**Log of Activities**

**Must record attendance at lectures and supervisions**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Date** | **Activity**  **incl. lectures & supervisions** | **Reason if not attend lecture or supervision** | **Weekly project progress.**  **How lecture/supervision was helpful to your project.** |
| 1 | 14/05/2024 | Lecture 1 |  | In this lecture we got the introduction about the project. The timeline of the project and choosing the dataset for the project |
| 2 | 21/05/2024 | Lecture 2 |  | Learned about the project and data management plan. |
| 3 | 28/05/2024 | Supervision 1 |  | Discussed about my interest in proposed projects “Traffic and charging station prediction for Electric vehicle optimization”. Discussed about my inability for finding the dataset. Supervisor proposed to takeup another project if I couldn’t find the dataset for EV. |
| 4 | 03/06/2024 | Lecture 3 |  | Learned about data ethics, UH policies regarding the datasets |
| 5 | 06/06/2024 | Supervision 2 |  | I couldn’t find the dataset for EV. I choose Brain tumor MRI dataset. But there are 4 people who chose the same dataset. So my supervisor proposed for me to look at prostate cancer dataset and I confirmed to take up as my project |
| 6 | 12/06/2024 | Supervision 3 |  | Today I gave the presentation for the project and data management plan as part of the 10% weightage of the project. After that I discussed my doubts regarding the dataset because of its complexity and my supervisor guided me well. |
| 7 | 17/06/2024 | Lecture 4 |  | Data Ethics part 2  Learned about the UK data protection ACT 2018, GDPR, Data Protection Impact Assessment, Data Storage & Security |
| 8 | 17/06/2024 | Supervision 5 |  | Discussed about the data pre-processing step, about understanding the data. |
| 9 | 25/06/2024 | Supervision 6 |  | Discussed and got some tips about the data Pre-processing step as the data is too clumsy. |
| 10 | 03/07/2024 | Assessment |  | Attended the Ethics quiz at the k110 lab |
| 11 | 08/07/2024 | Supervision 7 |  | Discussed the code that I have developed so far, including loading the images, extracting and labelling them, creating a model, dividing the labelled images into training and testing sets, running the model, and evaluating its accuracy. I also asked for guidance regarding the writing of the literature review and the methodology. |
| 12 |  |  |  |  |
| 13 |  |  |  |  |
| 14 |  |  |  |  |
| 15 |  |  |  |  |
| 16 |  |  |  |  |

**Log of GitHub Submissions**

**Record the versions of code and user documentation submitted on GitHub**

|  |  |  |
| --- | --- | --- |
| **Date** | **Filename and version submitted to GitHub** | **Description of code and/or documentation submitted (what has been added since the previous version).** |
| 06/07/2024 | Prostatexproject.ipynb | The prostate data set has been downloaded and uploaded in the google drive after that I loaded the dataset into the colab and tried to print images |
| 07/07/2024 | Prostatexproject.ipynb | inspecting the dataset. loaded the .csv files and inspected the rows and got all the t2 series images and saw the paths and the csv files for progressing the next step |
| 07/07/2024 | Prostatexproject.ipynb | I merged the findings and images CSV files into a single dataframe. I then extracted all the T2 image paths and created another dataframe by merging these image paths with the merged dataframe. |
| 08/07/2024 | Prostatexproject.ipynb | Extracted the images from the paths and labelled them as "cleansing true" or "cleansing false". Then, created a CNN model and trained it by dividing the labelled data into 70% for training and 30% for validation. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |