Reflective Journal Entries

Reflective journal identifies a person's working in project that what a person contributed to the project and what they learned through the process. Internship for me is the reason for reflective journal entries to see where I am lacking, what experience I gained throughout the internship progress, what I learnt regarding data science and if I had any difficulty understanding the process.

Using the formula D-I-E-P provided in the sample template, I will be discussing all week's objective learning and what task I implemented in this internship till .

Week 1

DESCRIBE

In Week 1, I got the chance to meet with my mentors in this internship which is Stephen CEO of ITIC Systems and Masoud Saifian who is a PHD candidate from MQU, and where I got to know about the introduction about the company and their dealings with clients and how they work in an organization, I got to know about high level idea of the project how we are going to proceed and with what timeline, then I got a chance to see the office and my working environment which was new to me. I learnt about data exploration and data analysis using various sources and tools, additionally I found the use of various tools to communicate online such as Microsoft meet, confluence that will help us in communicating with each other when we are at remote location.

INTERPRET

I gained new insights in working in an organization and working of an organization. Use of new tools for certain tasks gave me idea of flexibility when working between various departments, and data generation and collection of data from various sources gave me idea about mining big data and finding structure.

EVALUATE

After getting insights about organization, working, project I found this week to be productive for future as I was getting exposure of working in a company within a team and learning about the business model taught by CEO of company. Importance of teamwork using various communication is the key for successful team task which I got to know by analyzing company's employees. HR introduced us with all the policy while working as an intern and importance of it which helped me in learning basic idea of how to perform a task responsibly.

PLAN

As we progress further into weeks, I will be implementing all the learning from week 1 into rest of the internship and in future employment as most of the rules and regulations are

quite similar in organizations when working in an IT department. Learning new tools and techniques for better understanding of the concept will be my priority as these methods will help me in better coordination and communication in long term. The exposure from these learnings will reflect on my upcoming weeks as I will try to implement them with my own unique methods to get better results.

Week 2

DESCRIBE

In Week 2, I got the chance to learn about the project introduction, tools that we will be using to work on the project, work that have already been done on that project and lastly, I got to know about my role in the project which was as an intern. New tools were asked to learn to understand the basic concept of what I was doing with the data in an organizational setting. On other working day I was asked to generate data using generative AI tools like ChatGPT to test the dummy data, then later in that week I was asked to demonstrate what I have done throughout the weak in a presentation and what my ideas are for future tasks.

INTERPRET

I gained new insights in working with generative AI for commercial use and got to learn about vast amount of data that I needed to generate and work with, I learnt about Adobe products to use them on that same data to get some insights. For the first time I gained experience in giving presentation in office setting with all the outcomes from that week.

EVALUATE

After getting all the experience from that week I got to learn to communicate and work on my soft skills in which I was lacking, I got to learn about data security, governance and other important aspects that will help me in industry. Then I got to learn about the proper use of tools for a specific task which was ChatGPT in this case to generate the data. Overall, I gained experience working with a team with Stephen as our mentor and got to know about the details of my first commercial project.

PLAN

As we progress further into weeks, I will be dealing with data analysis for insights, UI design for the project in a working prototype, Machine learning Algorithms to determine the difference between fake essay data and real essay data, AI integration to check for any suspected cheating or contracted working and project management aspects to manage time, cost, and scope of the project to increase efficiency. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

Week 3

DESCRIBE

In Week 3, I got the chance to learn about Machine learning algorithms that will be used in the project and I was introduced with EssayGAN which was an essay data generation tool to generate dummy data to feed in that ML algorithm, then next working day I was asked to learn about the coding part of EssayGAN to know about the libraries that were used in that AI and to study them extensively to find any useful application in our project. The last working day of week 3 working day I was asked to display my ideas of using EssayGan into the project, and what libraries can be used in the generation process without any error, it was all presented via a team meeting and presentation in front of whole team.

INTERPRET

I gained knowledge about existing algorithms that used in a public project that we are using in our automated marking tool but in a different way to get any ideas from the existing technology. I gained exposure to commercial tools and coding in an project to understand the idea behind the code, new libraries was introduced to me to interpret and use in the machine learning algorithm and lastly, improved ideas about my soft skills while presenting and working with the team.

EVALUATE

After getting all the experience from that week I got to learn to explore current tools with algorithms which can be used in our project that shows. I increased my knowledge in IT domain and work with professionals to express my ideas and thinking behind the use of certain library in a machine learning algorithm, EssayGan was a completely new tool for me which was hard to learn initially but quite easy after some basic understanding.

PLAN

As we have already learnt about the project team, project information and generative AI in the project next step for me was to work with UI design for the project in a working prototype, Machine learning Algorithms to determine the difference between fake essay data and real essay data, AI integration to check for any suspected cheating or contracted working and project management aspects to manage time, cost and scope of the project to increase efficiency. Lastly,I will be looking for project outcome, result and suggestions if there are changes needed to be done.

Week 4

DESCRIBE

In Week 4, I got indulge in the machine learning feature extraction part where I was given the brief introduction about the feature and word extraction in the essay data to further analyze it for automated marking. Vector conversion of words into certain values was other task which was done in second day of working and then on the last working day of the week I was asked to do the same process which I was doing for 3 weeks that was to show my working, research and present my ideas for future works in that project.

INTERPRET

I gained knowledge various new feature extraction tools such as TF-IDF to convert words into vectors to feed into machine learning algorithms. I gained exposure in vectorization of words for efficient working with large amount of data using various techniques. Although I had knowledge about this matter of subject through my university but getting hands on experience in this field was relatively new for me and taught me about the depth of using machine learning for different tasks and integration of all the algorithms with one another.

EVALUATE

Week 4 was a hard week for me as there was plenty of research that needed to be done to start working on the algorithm. After an extensive reading and research, I got to know about the basic idea of why we are using vectorization and why we need feature extraction for our automated marking tool, after learning and presenting my learning in front of the team I am quite confident working with Tf-DIF IDF vectorization techniques in the future projects and rest of my internship program.

PLAN

As we have already learnt about the project team, project information and generative AI, feature extraction in essay data in the project next step for me was to work with UI design for the project in a working prototype, Machine learning Algorithms to determine the difference between fake essay data and real essay data, AI integration to check for any suspected cheating or contracted working and project management aspects to manage time, cost and scope of the project to increase efficiency. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

Week 5

DESCRIBE

In Week 5, I was given choice to either work with BERT which is a Bidirectional encoder to encode words into vectors using AI or to work in the making of UI design for the

project or frontend of the project. As everyone was choosing the BERT working, I challenged myself to learn about other aspects of the IT industry which was UI design for which had little to no knowledge about. Throughout the week I learnt about the basics of designing and implementing but with the help of Stephen at every step we collaborated and made a flowchart of working idea to implement it in our UI design. On the last working day of week 5 I explained each feature of UI Design that was done at that time to whole team and what are my future ideas that I will work on that UI in upcoming weeks.

INTERPRET

I gained knowledge a completely new domain which was related to my work, and which gives me flexibility to work with completely different tasks in future in an organization. All the reassert and analysis led to a prototype of the UI which was done while keeping in mind the actual idea of the project and simplification for others to understand what I was trying to implement in that UI. A simple yet successful UI design for first 2 web page was presented in the last working day of the week and all the learning worked as intended.

EVALUATE

Week 5 was a hard week for me as there was 2.5 working days and a completely new task was assigned to me which was never taught to me, nor I worked with it. I had to start from the beginning of UI designing as completely new tools and methods was required, and due to short period of time I worked and learnt about UI design in my free time as well before the team meeting to showcase them what I was working on and what can be done in the future of this project. A completely new idea was discovered as an addition to this UI and then in team meeting it was discussed whether to introduce a new feature to current project or not.

PLAN

As we have already learnt about the project team, project information and generative AI, feature extraction in essay data and UI Deign in the project next step for me was to work with Machine learning Algorithms to determine the difference between fake essay data and real essay data, AI integration to check for any suspected cheating or contracted working and project management aspects to manage time, cost and scope of the project to increase efficiency. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

Week 6

DESCRIBE

In Week 6, I was given introduction about the use of various AI algorithms and comparison between the different AI but the main AI algorithm which we will work with was Google BERT which is a bidirectional Transformer Encoder to encode and decode our data when used in our model, fine tuning of BERT was the task of the week to prepare BERT for different input such as

different languages and different structure. As an output BERT would be able to detect and encode any kind of language or input is the task that I was given with and to study its use and application in our model was to be conducted and on the last working day of the week I was asked to show my working throughout the week to show how it can be fine-tuned for our model.

INTERPRET

I gained knowledge about the basic structure and working of BERT model and other similar models to encode the words of an essay, we did comparison to find the most compatible model which can be used in our project, and it turned out to be BERT can be an ideal model if fine-tuned properly before the work. A successful knowledge and research were done and presented at the end of the week in the team meetings.

EVALUATE

Week 6 was a productive week as I gained knowledge about the new innovative tools that can be used in our current and future projects and this knowledge will help me in working in this subject with efficiency and accuracy. Fine-tuning or pre preparing BERT was a new experience and I got to learn about the basic concept of AI models which has a higher growth potential in field of data science.

PLAN

As we have already learnt about the project team, project information and generative AI, feature extraction in essay data and UI Deign and basic understanding of AI algorithms like BERT in the project next step for me was to work with, AI integration to check for any suspected cheating or contracted working and project management aspects to manage time, cost, and scope of the project to increase efficiency. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

Week 7

DESCRIBE

In Week 7, I did learn about the infrastructure of the 3 most important features of this project which is making an machine learning algorithm to generate a model which will score our dataset essay and the other two features are answer summarization which we did using commercial APIs and existing tool that does similar functions to it lastly, Question Generation is done by the software team to generate questions generated by the API in use.

INTERPRET

I got to learn about the initial test train model using the BBC dataset which was a sample dataset to check our model accuracy and then I learnt to tweak the model according to

our need that gave me new insight about the different libraries we used in this model. I learnt about the existing tools and how we can use the commercial API in our project.

EVALUATE

After completion of this week I successfully built a machine learning model from BBC News topic paper based on Classification for our model (https://github.com/dhrubasil/BBC-News-Classification) I learnt about different models and its working throughout the process and lastly I got to present my insights about the implementation of various models together as proposed in the flowchart[2].

Week 8

DESCRIBE

In Week 8, using the same functions and machine learning algorithms I was introduced with new data which was called ASAP dataset that consisted of marks of students from English exams and various E-Rater as well as exam marks from different domain experts, in order to predict weather the marks were correctly allocated on the summary we used Epoch training, Bert tokenizer and Bert model and the output results have been shown in the working outputs below.

INTERPRET

Initial steps as a junior data analyst was to identify any irregularities or any noise in in the data in any form and we found out that rater domain_1 had score from 0-60 which was beyond the limit of the specified model training so we tuned the data remove all the null columns and rows in order to increase the efficiency of the model and then took labels as from score 0 to 18 which was the most efficient score in this model and got about accuracy of 55% above.

EVALUATE

After completion of this week I successfully built a machine learning model from ASAP Dataset based on English exam, I learnt about different models and its working throughout the process, cleaning of a industry data was the most interesting part I took a lot of efforts and handwork to analyze the data in initial phase using graphs, structure and lastly I got to present my insights about the implementation of various models together as proposed in the flowchart.

PLAN

As we have already learnt about the project team, project information and generative AI, feature extraction in essay data and UI Deign and basic understanding of AI algorithms like BERT and lastly ML models our next step was to further implement commercial API's into our project, AI integration to check for any suspected cheating or contracted working and project management aspects to manage time, cost, and scope of the project to increase efficiency. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

WEEK 9

DESCRIBE

In Week 9, we started working on both BBC and ASAP data to improve the efficiency of our model by fine tuning the BERT model. As we got the initial results of both the models, we tend to look at the training side of the model where we changed various parameters, implemented more deep learning models with different learning rates, epoch trainings and others[2].

INTERPRET

Initial model did give us more loss in the training phase and but after getting the gist of the Bert tuning all those losses turned down with each epoch and with increased accuracy of the model which taught us about the importance of tuning a deep learning model with various constrains and also I got to know that it we cannot always get good results with all more learning rate or increased deep learning layers but it is about the dataset interpretation and what the model need according to the need of the user.

EVALUATE

This week was an important week for me as an individual in data science as I got to know about the small things that matter in the machine learning modelling and analysis process. As it was taught that the more the parameters the better, but I got to learn that it is all about data structure, model handling and fine tuning and extracting every bit of a result from the same model and by not changing the whole model.

PLAN

As we have already learnt about the project team, project information and generative AI, feature extraction in essay data and UI Deign and basic understanding of AI algorithms like BERT, and fine tuning of the machine learning model using industry data and lastly ML models our next step was to further implement commercial API's into our project, AI integration to check for any suspected cheating or contracted working and project

management aspects to manage time, cost, and scope of the project to increase efficiency. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

Week 10

DESCRIBE

In Week 10, for this week we were given with the corpus dataset which was a data that contained the English test score results of the candidates in an English exam in the UK (https://ilexir.co.uk/datasets/index.html) when looking at the steps we performed in the earlier data exploration they were data with good structure, but this was not the case of this data as there were 1244 files that was all sliced from years, different folder and was in very different extension in each folder that made us work on the dataset alone whole week to further process and feed it as an input to our machine learning Bert model [1].

INTERPRET

First encountered with the unstructured data was a different experience for me as I was working with mostly structured or semi structured data and that was easy to load, process and analyze but this week gave us all a challenging time and as a team we discussed what approach we can take to work on the data.

EVALUATE

Data preprocessing of this data was initially hard but after team effort and a lot of discussion we managed to create a single csv file to work with removing all the junk data and keeping the relevant fields for our ML implementation.

PLAN

As we have already learnt about the project team, project information and generative AI, feature extraction in essay data, UI Deign, basic understanding of AI algorithms like BERT, and fine tuning of the machine learning model using industry data and lastly ML models and preprocessing unstructured data. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

Week 11

DESCRIBE

In Week 11, company meeting, and farewell party was organized to celebrate the 11 weeks of work in ITIC where we were invited by Stephen for an afternoon lunch in a nice place with Masoud and Jennifer. Where we got many insights about the Australian culture, working, and job placement scenario and was asked to work on our CV for future opportunities.

INTERPRET

As it was a short break for the week we got to know about the importance of connection, speaking skills with co-workers, soft skills in an organization and mostly the insights of working in IT in Australia.

EVALUATE

Insights from all three mentors, Stephen, Masoud and Yennifer it was truly a motivation to work hard to learn new techniques to achieve success in this field of industry. The rewarding part of this week was to discuss things outside the work about the challenges we might face, the experience from my mentors and others gave me an insight about how I can work on my skills and enter the IT domain.

PLAN

As we have already learnt about the project team, project information and generative AI, feature extraction in essay data, UI Deign, basic understanding of AI algorithms like BERT, and fine tuning of the machine learning model using industry data and lastly ML models and preprocessing unstructured data. Lastly, I will be looking for project outcome, result, and suggestions if there are changes needed to be done.

Week 12

DESCRIBE

In Week 12, we worked on the clean corpus data set for exploration and checked if there is further cleaning required or not and then we loaded that data into our machine learning algorithm with Bert then we did some visual charts such as bar graphs and line graphs to see the data consistency and frequency. Then we fed our tokenized data to the model and 40 epoch trainings to get the result and present it on the last working day of the week.

INTERPRET

Overall learning experience in this week was to interpret the data for input for pre-defined model was quite difficult and was time consuming. Label setting and other fine tuning of the model was the next step that was followed and in the last we got the accuracy of the model to see how it was performing.

EVALUATE

All the model training and fine tuning was a critical step when dealing with such dataset with inconsistencies and overall, it was a successful week as we got to know about the handling such dataset. Rest of the steps was same as for the ASAP dataset and BBC dataset for model training and testing.

PLAN

Before the last week we have already set up a high level idea for our project, explored other tools that are currently in the market, set up a journey map or flowchart to follow steps, researched about the code snippet related to machine learning, learnt about BERT its operations and working, UI design Including all the features that will be required in the web page, Bert model initialization, pre trainings and training models on different datasets. Now the very last task was to save all the progress in the repository for future interns to work on and to move this project forward.

Week 13

DESCRIBE

In Week 13, it was all about the feedbacks how we did we progressed in the internship and project, saving all the progress of the work we did throughout the internship into a repository, making reports of all the outputs for future interns as a reference.

INTERPRET

This week task was about to save all the checkpoints for the future students and to reflect on our own performance throughout the internship. And lastly making a final report on the project and new future ideas that can be added into this project.

EVALUATE

All the progress was saved and mailed to our supervisor, code stubs and snippets was saved in the LMS that is easily available to everyone and further flowchart was also discussed to plan for future students in order to work efficiently.

PLAN

Before the last week we have already set up a high level idea for our project, explored other tools that are currently in the market, set up a journey map or flowchart to follow steps, researched about the code snippet related to machine learning, learnt about BERT its operations and working, UI design Including all the features that will be required in the web page, Bert model initialization, pre trainings and training models on different datasets. All the work that was intended for this batch is now completed in this week and all the workflow was saved for all the future interns to work on [8].

Work Samples

UI DESIGN IDEAS FOR AUTOMATED MARKING TOOL ASSESSMENT

As an Intern I was given with two task choices weather to work on BERT research paper and code analysis or present a design of the UI that will be used on the web, I choose to work on the UI Design as that as new knowledge base for me and I worked on this UI for about 2 weeks before moving on to my next task.

A detailed description of UI task with output picture are shown below.