| **1. PROJECT DETAILS** | | | |
| --- | --- | --- | --- |
| **Student Name** | **RYAN KHO YUEN THIAN** | | |
| **Programme**  **(*Circled The Option*)** | **RST / REI / RSD / RIS / RSW / RDS / RIT** | | |
| **Project Title** | **A Study on whether Automated Machine Learning or Generative AI can**  **replace the Human Data Scientist** | | |
| **Supervisor Name** | **Ms FATIN IZZATI BINTI RAMLI** | | |
| **Category**  **(*Circled The Option*)** | **Development / Research** | **START OF FYP**  **(*Circled The Option*)** | **20 …… FEB/MAY/SEPT** |

**IMPORTANT NOTES :**

1. Students shall log the details of the progress meeting in this document.
2. Please duplicate the blank form if the number of meetings exceeds the forms provided in this document.
3. It is the student's responsibility to keep this document updated. All suggestions, ideas, comments, and instructions given by the supervisor need to be recorded in the log book and to be complied with in the project.
4. The student must have a regular meeting with the supervisor. **A minimum of 6 meetings are required throughout the semester.**
5. This document shall serve as the evidence for final assessment and grading.
6. This document shall be submitted with the FYP 1 report at the end of the semester.
7. The faculty reserves the right not to accept the final report for examination if this document is incomplete or not submitted on the stipulated dateline.

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** | **12th July 2024** | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | * Wrote almost the full Machine Learning Code for the Human Data Scientist Approach * Wrote almost the full Machine Learning Code for the AutoML Approach (H2O AutoML) * Learnt how to use FastAPI * Created the User Interface for the Human Data Scientist Approach using Streamlit | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | * At end of week 12, present the application. * Submit everything by Friday of week 12. * Implementation & Testing: include code & test cases * Deployment: can be either web based or local * Subsequent meetings with supervisor to be individual & face-to-face. Each meeting is 15 mins long but if more time is needed, it can be allowed * Chapter 5 - show modelling codes + explain how to do testing * Appendix - show only the necessary codes, which were not mentioned in Chapter 5 like data preprocessing * Chapter 6 - deployment into the real world environment. If deployed into a web server, need to mention how it is done? Need to include words + diagrams * No need user manual * Can use either Harvard or APA referencing style | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | * Show progress of system | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** | **RyanK** | | | **Supervisor Signature :** | | | | | | | ***Fatin*** | |  |
|  | | |  |  | | | | |  | | | | | |  | |

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** | **15th July 2024** | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | * Progress for Human Data Scientist Approach is still the same. * Selected and deployed a H2O AutoML Model using FastAPI & Streamlit. The app for AutoML approach can now make prediction but I have not displayed the Shap Summary Plot along with the prediction * Haven’t started on Generative AI approach & haven’t registered a ChatGPT Plus account yet | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | * good progress * complete the summary of outcome for AutoML * continue working on all 3 approaches * measure performance on sample data for AutoML approach * compare all 3 approaches later on | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | * work on all 3 approaches * display the Shap Summary Plot with AutoML prediction on Streamlit * improve the Keras model and create app for Human Data Scientist approach * measure the selected AutoML model’s performance on sample data | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** | **RyanK** | | | **Supervisor Signature :** | | | | | | | ***Fatin*** | |  |
|  | | |  |  | | | | |  | | | | | |  | |

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** | **25/7/2024** | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | * Fixed the configuration of the Keras Classifier model and made it to focus on maximising AUC rather than accuracy * Displayed the performance metrics for the AutoML model * Improved the efficiency of the FastAPI app for AutoML app * Managed to display the SHAP summary plot in AutoML app * Learnt the capabilities of ChatGPT Data Analyst via ChatGPT 4o model * Managed to launch local tracking server for MLFlow and use its UI to track the various ML runs and metrics | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | * good progress * need to explain the values on the y axis of the SHAP summary plot * need to interpret the values displayed for each performance metric of the AutoML model * deploy the app for Human Data Scientist approach * research into Generative AI approach | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | * need to explain the values on the y axis of the SHAP summary plot * need to interpret the values displayed for each performance metric of the AutoML model * deploy the app for Human Data Scientist approach * research into Generative AI approach | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** | **RyanK** | | | **Supervisor Signature :** | | | | | | | ***Fatin*** | |  |
|  | | |  |  | | | | |  | | | | | |  | |

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** | **29/7/2024** | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | * Selected KerasClassifier as the best model * Saved and Deployed KerasClassifier using a different approach * Almost completed the application for the Human Data Scientist Approach. Able to make prediction but currently unable to generate the explanation plot. * Made changes to the AutoML notebooks by measuring the performance of the top 5 AutoML models (focused on Recall only). Chose GBM Grid model as the best AutoML model * Did some documentation for Chapter 6 System Deployment | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | * Good progress * The KerasClassifier model deployed in the Human Data Scientist application is able to make a reasonable prediction. For eg, if one enters values that indicate a risky loan applicant, the application will indicate that the applicant is likely to default. * When I am done with all 3 approaches, I should enter the same set of input data for all 3 to determine whether they return the same predictions. * Chapter 6: write about the method of deployment, how big the server is, how you are able to link it to the prediction model, how you upload the prediction model, etc. | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | * Start work on the Generative AI approach * Generate the explanation plot for the Human Data Scientist approach | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** | **RyanK** | | | **Supervisor Signature :** | | | | | | | ***Fatin*** | |  |
|  | | |  |  | | | | |  | | | | | |  | |

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** | **6/8/2024** | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | * Obtained a ChatGPT Plus account via subscription * Uploaded German Credit dataset to ChatGPT Data Analyst (CDA) and prompted it to generate and execute codes necessary for Machine Learning (from Data Understanding to Deployment) * CDA was unable to execute code from GridSearch onwards and unable to display what the Streamlit app would look like * Still in the process of debugging the Streamlit & FastAPI code generated by CDA. | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | * Continue working on the Generative AI implementation * In FYP report, highlight the weaknesses of ChatGPT/Generative AI and indicate where the Human must be involved (to improve coding) * Compare the UI created by GenerativeAI with the other 2 approaches, mention in what aspects GenerativeAI can improve and use the ideas from GenerativeAI to improve the manual one | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | * Complete the implementation & testing for the GenerativeAI approach | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** | **RyanK** | | | **Supervisor Signature :** | | | | | | |  | |  |
|  | | |  |  | | | | |  | | | | | |  | |

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** | **12/8/2024** | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | * Completed the implementation & testing for the Generative AI approach * Wrote subheadings & descriptions in the Generative AI jupyter notebook | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | * In the documentation, mention that I had to prompt ChatGPT several times to get good outcome * In the documentation, mention the differences between ChatGPT approach & other approaches * For Chapter 5, show the important parts of the code from beginning to end (put less imp parts in appendix). Should include all stages in the machine learning life cycle. * For Chapter 5 test cases, mention 10-15 cases (see whether you get the expected prediction based on values entered) * Deploying the system locally is acceptable. | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | * Do some write-up for Chapter 5 * Finalise the coding | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** | **RyanK** | | | **Supervisor Signature :** | | | | | | |  | |  |
|  | | |  |  | | | | |  | | | | | |  | |

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** |  | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | <WRITE HERE> | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | <WRITE HERE> | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | <WRITE HERE> | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** |  | | | **Supervisor Signature :** | | | | | | |  | |  |
|  | | |  |  | | | | |  | | | | | |  | |

| **PROGRESS RECORD** | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | | | **:** |  | | | | **MEETING** | | | | **:** | **1 / 2 / 3 / 4 / 5 / 6 / 7 / 8 / ……….** | | | |
|  |  |  | | | |  | | | |  |  | | | | | |
| **Progress Completed After Previous Meeting**  **(*To be completed before meeting*)** | | | **:** |  | <WRITE HERE> | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Supervisor Comments / Suggestion** | | | **:** |  | <WRITE HERE> | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Next Meeting Expected Progress** | | | **:** |  | <WRITE HERE> | | | | | | | | | | | |
|  | | |  |  | |  | | | |  |  | | | | | |
| **Student Signature** | | | **:** |  | | | **Supervisor Signature :** | | | | | | |  | |  |
|  | | |  |  | | | | |  | | | | | |  | |