**PRJ3213: Capstone Project 2**

**Activity Log**

**Analyzing the Factors Affecting Customer Churn and Predicting Customer Churn for a Telecommunication Industry**

**Chin Wai Chung**

**18121376**

**Bachelor of Information System (Data Analytics) (BSDA)**

**Sunway University**

**Supervisor:**

**Dr. Melody Tan Shi Ai**

# Table of Contents

# 1. Gantt Chart and Work Plan

Building upon the Capstone 2 schedule, section 1 contains Gantt charts mapping out the timeline for executing Capstone Project 2 spanning weeks 1-14. The pair of charts outline the phases, milestones, and week-by-week breakdown of activities to complete the final year analytics project.
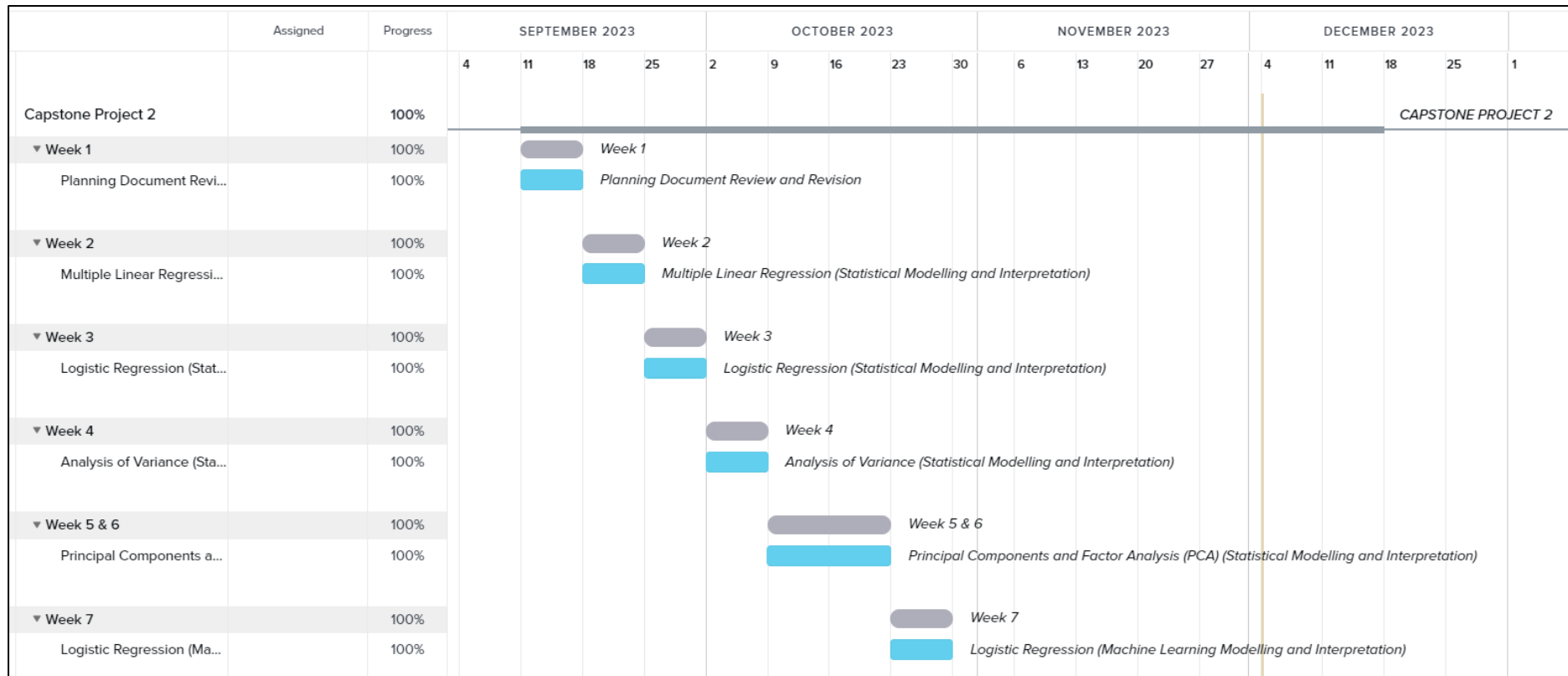


**Figure 1.1**: Gantt Chart for Capstone Project  2: Final Year Project (Week 1 – 7)
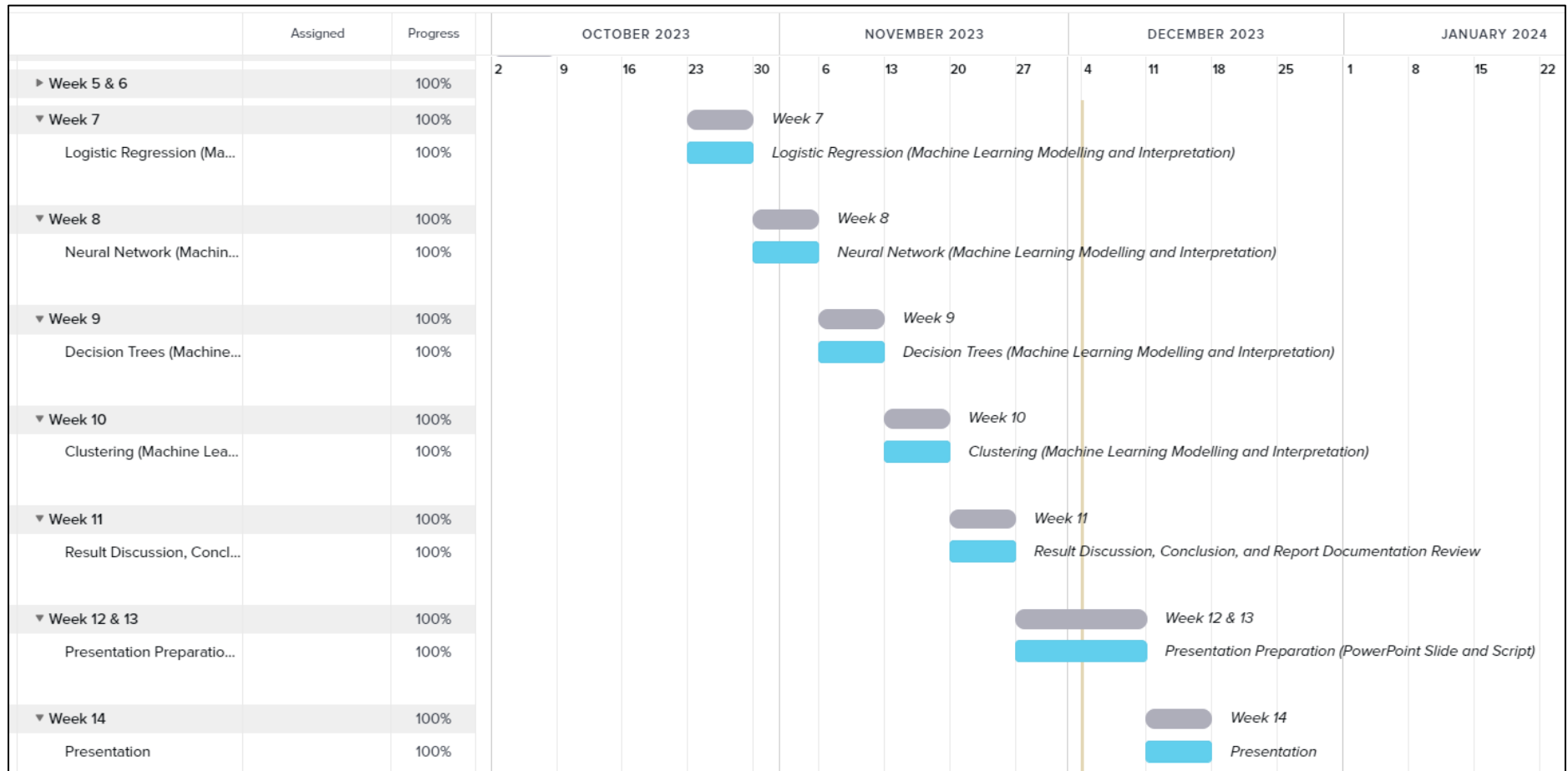
**Figure 1.2**: Gantt Chart for Capstone Project 2: Final Year Project (Week 8 – 14)

## 1.1 Activities for Capstone Project 2: Final Year Project

**Week 1 - Project Planning Review**

The first week is dedicated to thoroughly reviewing the project planning document, methodology, and proposed timeline to identify any required changes or improvements before execution gets underway. It's critical to ensure the research plan is comprehensively optimized at this point, since major deviations later can cause significant rework and delays. Key tasks involve re-examining the problem statement, objectives, background literature, data sources, methodology selections, and Gantt chart schedule. Meetings with the supervisor help identify blind spots. Refinements are made to strengthen the plan where prudent. This upfront planning review reduces downstream risks and inefficiency.

**Weeks 2-5 - Statistical Modelling**

Weeks 2 through 5 focus on developing the statistical modelling component of the project methodology. Multiple regression is modelled in Week 2 to quantify the isolated effects of churn drivers like service quality and contract tenure while controlling other factors. Week 3 centers on implementing logistic regression to estimate churn probabilities based on customer attributes like demographics, tenure, usage metrics etc. Week 4 shifts to ANOVA analysis to detect significant differences in churn rates across customer segments defined by attributes like age brackets and plan types. Finally in Week 5, principal component analysis and factor analysis techniques are leveraged to consolidate correlated variables and distil the numerous detailed predictors into core underlying latent factors that are better suited for parsimonious modelling. Together, these 4 weeks of statistical modelling supply a crucial explanatory understanding of churn driver relationships and effects to complement the machine learning predictive modelling.

**Weeks 6 - Design Refinement**

With the foundational statistical analysis complete, Weeks97 are allocated as a buffer period to refine the remaining project design based on insights gained so far. Key tasks involve reviewing the statistical findings to finalize which data variables and feature transformations should feed into upcoming machine learning models. The methodology for churn customer segmentation, validation techniques, and prototype design are also optimized based on learnings. This lookback review allows course-correcting

the remaining phases to maximize their effectiveness. The 2-week duration provides flexibility to iterate on improvements.

**Weeks 7-10 - Machine Learning Modelling**

Weeks 7 through 10 shifts focus to developing machine learning models for churn prediction, leveraging the prepared data and insights from statistical analysis. Week 7 focuses on implementing logistic regression as a baseline classifier before exploring more advanced nonlinear techniques. Week 8 examines neural networks to uncover subtle patterns missed by logistic regression. Decision trees are modelled in Week 9 to deliver churn segmentation profiles. Together, these 3 weeks apply diverse machine learning paradigms to accurately predict churn while generating intuitive explanations. While week 10 will be focusing on creating a cluster model to examine the relationship between each of the variables.

**Weeks 11 - Results Discussion and Report Review**

With the modelling complete, Weeks 11 are concentrated on synthesizing results, deriving conclusions, and reviewing report drafts. Discussion meetings with supervisors help interpret key takeaways and limitations. Conclusions are formulated based on objectives. Report drafts are reviewed to ensure accurate documentation of the end-to-end technical methodology and business insights gained. These 2 weeks allow time for thoughtful analysis of the findings and thorough report refinement.

**Week 12-13 - Presentation Preparation**

Week 12 to 13 focuses on preparing the presentation that will convey the capstone project's core concepts, findings, recommendations, and contributions to stakeholders. Visualizations are created to clearly communicate concepts. Talking points are refined to emphasize key takeaways within the time limit. Practice rounds verify the delivery timing and flow. This presentation preparation week completes the project, allowing the student to clearly articulate the business value delivered through this analytical initiative.

In summary, the Gantt charts demonstrate a comprehensive 13-week schedule to execute the data science capstone spanning statistical analytics, machine learning modelling, results analysis, and presentation. The timeline balances efficiency with diligent phase durations to ensure prudent project management.

## 2. Timeline and Activity

**Table 1.1:** Timeline and Activity

| Capstone Project 1 | | |
|---|---|---|
| **Week** | **Activity** | **Description** |
| 1 | − Review Project Planning<br>− Methodology Assessment<br>− Timeline Evaluation<br>− Data Source Examination,<br>− Meetings with Supervisor<br>− Plan Refinement<br>− Risk Reduction Analysis<br>− Document Changes | This week primarily involves a comprehensive review of the project planning document, methodology, and proposed timeline to ascertain their alignment with project objectives and reduce potential downstream risks. The review entails scrutinizing the problem statement, objectives, background literature, and selected methodology for optimization. The Gantt chart schedule undergoes assessment to identify bottlenecks and dependencies, while data sources are carefully validated for reliability and sufficiency. Meetings with the supervisor help detect blind spots, refine the plan, and gather feedback for enhancements. Subsequently, necessary refinements are made, and potential risks are identified and mitigated to fortify the plan. All changes, improvements, and documentation are meticulously recorded for future reference and tracking purposes. |
| 2<br><br>3<br><br>4<br><br>5 | − Week 2 - Multiple Regression Modelling<br>− Week 3 - Logistic Regression Implementation<br>− Week 4 - ANOVA Analysis<br>− Week 5 - Principal Component | Over the span of four weeks, the focus is on statistical modeling techniques vital for understanding churn drivers and behaviors. Week 2 involves multiple regression modeling to quantify the isolated effects of churn drivers like service quality and contract tenure. In Week 3, logistic regression is implemented to estimate churn probabilities based on various customer attributes. Week 4 shifts to ANOVA analysis, identifying significant differences in churn rates across distinct customer segments. Lastly, in Week 5, principal component analysis (PCA) and factor analysis are |

| Capstone Project 1 | | |
|---|---|---|
| **Week** | **Activity** | **Description** |
| | Analysis & Factor Analysis. | employed to consolidate variables into core underlying factors, aiming for a simplified set of predictors better suited for subsequent modeling. Together, these four weeks of statistical modeling aim to provide comprehensive insights into churn behavior to complement the upcoming machine learning predictive modeling phase. |
| 6 | − Weeks 6: Design Refinement | Utilize Weeks 6-7 as a buffer period dedicated to refining the project design based on insights obtained from foundational statistical analysis. Review the statistical findings comprehensively to finalize the selection of data variables and feature transformations that will be utilized in upcoming machine learning models. Optimize the methodology for churn customer segmentation, validation techniques, and prototype design based on the insights gained from the statistical analysis. This period allows for a retrospective review to fine-tune and course-correct the remaining project phases for enhanced effectiveness. Leverage the 2-week duration to iterate on improvements and adjustments for an optimized project design before progressing to subsequent phases. |
| 7 8 9 10 | − Week 7 - Logistic Regression Implementation − Week 8 - Neural Network Exploration − Week 9 - Decision Tree Modelling | During Weeks 7-10, the focus shifts towards developing machine learning models for churn prediction, capitalizing on prepared data and insights derived from statistical analysis. Week 7 initiates with implementing logistic regression as a baseline classifier to establish a foundational predictive model. In Week 8, neural networks are explored to uncover nuanced patterns that might have been missed by logistic regression, aiming for more intricate insights. |

| Capstone Project 1 | | |
|---|---|---|
| **Week** | **Activity** | **Description** |
|  | − Week 10 - Cluster Model Creation | Week 9 centers on modeling decision trees, offering churn segmentation profiles for deeper understanding and intuitive explanations. Additionally, Week 10 concentrates on creating a cluster model to investigate the relationships between variables, providing further insights into customer behavior and churn dynamics. These four weeks encompass diverse machine learning paradigms to predict churn accurately while facilitating intuitive interpretations and exploratory analysis. |
| 11 | − Results Synthesis<br>− Conclusions Formulation<br>− Report Draft Review<br>− Discussion Meetings with Supervisors | Week 11 is dedicated to synthesizing the modeling results, deriving meaningful conclusions, and conducting a comprehensive review of report drafts. The focus is on synthesizing the outcomes obtained from the completed modeling phases, interpreting key takeaways, and acknowledging any limitations discovered during the process. The formulation of conclusions is driven by the alignment with the initial project objectives. This week also involves extensive review sessions of report drafts to ensure accurate documentation of the end-to-end technical methodology applied throughout the project and to capture essential business insights gained from the analysis. Discussion meetings with supervisors are crucial during this phase to gain insights, interpret findings, and address any concerns or queries that might arise. These two weeks provide ample time for thoughtful analysis of the findings and meticulous refinement of the report to ensure its accuracy and completeness. |

| Capstone Project 1 | | |
|---|---|---|
| **Week** | **Activity** | **Description** |
| 12<br><br>13 | – Presentation Content Creation<br><br>– Visualizations Development<br><br>– Refinement of Talking Points<br><br>– Practice Sessions | Weeks 12-13 are dedicated to the meticulous preparation of the presentation that aims to effectively communicate the core concepts, findings, recommendations, and contributions of the capstone project to stakeholders. This period involves the creation of visualizations and graphical representations to vividly convey complex concepts and analytical insights. Refinement of talking points ensures the emphasis on key takeaways within the designated presentation timeframe. Practice rounds are conducted to verify the timing and flow of the delivery, ensuring a polished and coherent presentation. This phase serves as the culmination of the project, allowing the student to articulate and demonstrate the business value derived from this analytical initiative to stakeholders. In summary, these two weeks represent a critical phase in the project, aligning with the comprehensive 13-week schedule encompassing statistical analytics, machine learning modeling, results analysis, and the final presentation. The timeline strikes a balance between efficiency and thoroughness, reflecting prudent project management practices. |

### 3. Supervisor Meeting Record

This report compiles the supervision meeting records between student Chin Wai Chung and supervisor Dr. Melody Tan Shi Ai over the course of the project from 22/05/2023 to 31/07/2023.

## SCHOOL OF ENGINEERING AND TECHNOLOGY

### DEPARTMENT OF COMPUTING AND INFORMATION SYSTEMS

# SUPERVISION MEETING RECORD 1

Date: 20/09/2023

Time: 3.15 PM

Student: Chin Wai Chung

Supervisor: Dr Melody Tan Shi Ai

**Items discussed this meeting:**

1. Project Timeline

   - Agreed to complete the modelling process by week 10.
   - Emphasized the importance of adhering to the deadline for modelling tasks.

2. Dataset Check

   - Reminder to review the dataset thoroughly before initiating the modelling process.
   - Ensure data quality, completeness, and relevance for accurate modelling.

**Work for the coming meeting:**

1. Dataset Evaluation

   - Conduct a comprehensive review of the dataset.
   - Identify any missing, erroneous, or inconsistent data that may impact modelling.

2. Initial Modelling Preparation

   - Begin organizing necessary tools and resources for the modelling phase.
   - Prepare an outline or strategy for the modelling approach to ensure efficiency.

3. Documenting Modelling Plan

   - Draft a detailed plan outlining the steps and methodologies for the modelling process.
   - Include milestones, benchmarks, and potential challenges for reference.

4. Preparation for Progress Update

   - Gather relevant information or progress updates to present in the next meeting.
   - Prepare to discuss any potential roadblocks or assistance required for successful modelling.

**Supervisor's Signature**                    **Student's Signature**

Dr Melody Tan (7 Dec 2023)                    Chin

---------------------------------------------                    ---------------------------------------------

# SUPERVISION MEETING RECORD 2

Date: 04/10/2023

Time: 3.00 PM

Student: Chin Wai Chung

Supervisor: Dr Melody Tan Shi Ai

**Items discussed this meeting:**

1. Current Progress Update

    - Completed the second statistical model and progressing towards initiating the third model.
    - Continuing efforts toward enhancing the modelling processes.

2. Dataset Modification

    - Decision to change the dataset due to challenges faced with the previous proposed dataset containing entirely categorical data.
    - Identified difficulties in modelling with the previous dataset structure.

3. Tool Recommendation

    - Received a suggestion to utilize SAS Viya due to encountered issues with SAS Miner.
    - Acknowledged the need for exploring SAS Viya functionalities for seamless modelling.

**Work for the coming meeting:**

1. Dataset Transition

    - Identify and acquire the new dataset suitable for the modelling process.
    - Ensure the new dataset encompasses a balanced mix of categorical and other relevant data types for effective modelling.

2. SAS Viya Familiarization

- Begin familiarizing with SAS Viya functionalities and features.
- Explore resources or tutorials to understand its capabilities in statistical modelling.

3. Third Model Preparation

- Continue working on conceptualizing and preparing for the third statistical model.
- Plan the approach considering insights gained from the previous models and the updated dataset.

4. Troubleshooting SAS Miner

- Attempt to resolve the issues encountered with SAS Miner if applicable.
- Seek guidance or technical support to rectify any persisting problems with the tool.

**Supervisor's Signature**                                    **Student's Signature**

Dr Melody Tan (7 Dec 2023)                                    Chin

---------------------------------------------                                    ---------------------------------------------

# SUPERVISION MEETING RECORD 3

Date: 18/10/2023

Time: 3.30 PM

Student: Chin Wai Chung

Supervisor: Dr Melody Tan Shi Ai

**Items discussed this meeting:**

1. Progress Update - Statistical Modelling

   - Progressed significantly, reaching 80% completion of the 4th statistical model.
   - Estimated completion within the next two days, enabling the initiation of the final statistical model.

2. Upcoming Tasks

   - Plan to commence work on the last statistical model after completing the ongoing 4th model.
   - Intending to start the machine learning model development in the following week.

**Work for the coming meeting:**

1. Final Statistical Model Completion

   - Complete the 4th statistical model within the specified timeframe.
   - Initiate groundwork for the last statistical model, considering insights gathered from previous models.

2. Machine Learning Model Preparation

   - Gather necessary resources and tools required for machine learning model development.
   - Outline the initial steps and approach for the machine learning model based on project requirements.

3. Documentation and Progress Tracking

- Document progress made in completing the 4th statistical model.
- Create a progress tracking system or update existing documentation to monitor the advancements in the modelling phase.

4. Resource Preparation

- Ensure availability of datasets, software, and any additional resources needed for the machine learning model development.
- Confirm access to required platforms or tools for seamless work execution.

**Supervisor's Signature**                                    **Student's Signature**

*Dr Melody Tan (7 Dec 2023)*                          *Chin*

-------------------------------------------                          -------------------------------------------

# SUPERVISION MEETING RECORD 4

Date: 30/10/2023

Time: 08.00 AM

Student: Chin Wai Chung

Supervisor: Dr Melody Tan Shi Ai

**Items discussed this meeting:**

1.  Progress Update - Statistical Modelling

    - Successfully completed the 5th statistical model along with interpretation.
    - Transitioning towards commencing the machine learning modelling phase.

2.  Challenges Faced with SAS Viya

    - Encountered difficulties with data import in SAS Viya during the initial stages of transitioning to machine learning modelling.
    - Received guidelines to resolve the data import issue.

**Work for the coming meeting:**

1.  Machine Learning Modelling Initiation

    - Commence the groundwork for machine learning modelling after completing the statistical modelling phase.
    - Prepare the framework or plan for approaching the machine learning tasks.

2.  Resolution of Data Import Issue

    - Implement the guidelines received to address the data import problem in SAS Viya.
    - Ensure successful data import for a seamless transition into machine learning tasks.

3.  Resource Evaluation and Preparation

    - Assess the resources required for machine learning modelling, including datasets, tools, and software.
    - Confirm availability and access to necessary platforms or technologies for effective work execution.

4.  Documentation and Reporting

    - Document insights gained from the completed statistical models and their interpretations for future reference.
    - Prepare reports or summaries highlighting key findings from the statistical modelling phase.

**Supervisor's Signature**

Dr Melody Tan (7 Dec 2023)

--------------------------------------------

**Student's Signature**

Chin

--------------------------------------------

# SUPERVISION MEETING RECORD 5

Date: 15/11/2023

Time: 02.00 PM

Student: Chin Wai Chung

Supervisor: Dr Melody Tan Shi Ai

**Items discussed this meeting:**

1. Progress Update - Modelling Phase

   - Successfully completed the Logistic Regression and Decision Tree models.
   - On track to finish the remaining model by Tuesday.

2. Additional Model Consideration

   - Sought suggestions for an alternative to the originally planned Association Rule model due to its unsuitability for the project.
   - Received advice that utilizing four models, including Logistic Regression and Decision Tree, is considered adequate for the project's objectives.

**Work for the coming meeting:**

1. Finalize Pending Model

   - Complete the pending model by the specified deadline (Tuesday).
   - Ensure a comprehensive review and validation of the model's outcomes.

2. Evaluation of Model Suitability

   - Revisit the project requirements and objectives to confirm that the chosen models align with the project's needs.

- Document reasons behind the exclusion of the Association Rule model and ensure the adequacy of the chosen models.

3. Model Comparison and Analysis

- Prepare a detailed comparison among the completed models, highlighting their strengths, weaknesses, and suitability for the project.
- Gather insights and recommendations based on the analysis for the project's next steps or potential improvements.

**Supervisor's Signature**                                    **Student's Signature**

Dr Melody Tan (7 Dec 2023)                                    Chin

---------------------------------------------                        ---------------------------------------------

# SUPERVISION MEETING RECORD 6

Date: 29/11/2023

Time: 03.30 PM

Student: Chin Wai Chung

Supervisor: Dr Melody Tan Shi Ai

**Items discussed this meeting:**

1. CP2 Draft Review

   - Completed the review of CP2 draft and received suggestions for further improvements.
   - Acknowledged the need for enhancements based on the provided feedback.

2. Final Conclusion Clarification

   - Double-checked the approach for crafting the final conclusion of the CP2 report to ensure clarity and completeness.
   - Confirmed the specific elements or key points to be included in the concluding section.

3. Activity Log Discussion

   - Discussed the structure and content of the activity log required for the project.
   - Clarified any uncertainties regarding what information to include and how to format the activity log effectively.

**Work for the coming meeting:**

1. CP2 Draft Revision

   - Implement suggested improvements and enhancements to the CP2 draft.
   - Ensure clarity, coherence, and completeness in all sections, incorporating the received feedback.

2.  Final Conclusion Composition

    - Craft the final conclusion section of the CP2 report incorporating confirmed key points or elements discussed during the meeting.
    - Ensure the conclusion encapsulates the project's key findings, outcomes, and potential recommendations.

3.  Activity Log Preparation

    - Create a structured format for the activity log based on the discussed requirements.
    - Start documenting project-related activities, ensuring the log captures essential details accurately.

4.  Quality Check and Review

    - Conduct a thorough quality check of the revised CP2 draft, final conclusion, and initial activity log entries.
    - Review for coherence, accuracy, and adherence to project guidelines before submission.

**Supervisor's Signature**                    **Student's Signature**

*Dr Melody Tan (7 Dec 2023)*                    *Chin*

-------------------------------------------                    -------------------------------------------