



****

**Project Consultancy and Practicum**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

DataMicron Systems Sdn Bhd

Practicum Logbook submitted in partial fulfilment of the requirement for

**Masters of Science (Data Science and Analytics)**

For

***Dr. Nasuha Lee Abdullah***

***Dr. Chew XinYing***

By

**Muhammad Azzubair bin Azeman**

**P-COM0019/19**

Session

**Semester 2020 / 2021**

|  |  |  |  |
| --- | --- | --- | --- |
| **DATE** | **WORK DONE** | **CONTACT HOURS** | **REMARKS** |
| 12/10/2020 | Meeting with Client together with mentor. | 2 hours | Client presented their problem statement, current methodology, and solution expectations. In addition, they shared their dataset.  Problem Statement:  - Inaccurate PCI Forecasting  Solution Expectation:  - More Accurate PCI Forecasting using Machine Learning |
| 12/10/2020 | Internal Meeting with Mentor and colleagues | 2 hours | Mentor discussed strategies to meet up clients’ expectations. Have to find additional variables to enrich given dataset and more accurate PCI forecasting. |
| 13/10/2020 | Researched on additional variables referring on published papers | 7.75 hours | Found 2 additional variables based on published papers: |
| 14/10/2020 | Presented to Mentor on researched additional variables. | 1 hour | Mentor approved on the additional attributes and decided to proceed with data science tasks:  - Additional Data Collection  - Data Modelling  - Future PCI Forecasting |
| 14/10/2020 | Collected Additional  Variables | 3 hours | Searched additional variables in official data sources such as:  - Department of Statistics Malaysia (DOSM)  - Bank Negara Malaysia (BNM) |
| 14/10/2020 | Data Preprocessing | 0.25 hours | Filter Dataset  - Discarded invalid values (zeros)  - Omitted invalid attributes  Transform Dataset  - Aggregated values from daily to quarterly values |
| 15/10/2020 | Data Modelling | 36 hours | Modelling Dataset  - Researched Modelling python codes for Decision Tree, Random Forest, Support Vector Machine, Neural Network, and Extreme Gradient Boosting.  - Evaluated the Models |
| 19/10/2020 | Model Optimisation | 36 hours | Compared Model Performance  - Compared before and after Feature Selection  - Compared before and After tuning Horizon size  - Compared before and after combining Feature Selection and Tuning Horizon Size combination  - Selected top three models |
| 24/10/2020 | Model Forecasting | 24 hours | Forecasted 1 period in future  - Using the top three performing models, new future values were generated and input into the models to predict for future PCI.  - Models were tuned to achieve better performing forecasting accuracy by comparing with historical PCI values  - The best forecasting hyperparameters were set to forecast future PCI values |
| 27/10/2020 | Modelling and Forecasting Presentation to mentor | 2 hours | Mentor approves on the methodology and outcomes. The forecasting values will be evaluated by comparing with published PCI values that will be released by November 2020. |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |
| --- | --- |
| Remarks by Mentor |  |

Mentor’s Signature: ........................................................................ Date: ..................................