**Summarized Meeting Minutes – Group 07**

**Project Planning:**

* Prepared project plan for Capstone-2.

**Data Integration:**

* New column to the TTC dataset to integrate unstructured data was not added due to lack of resources to handle unstructured data sources on web like tweets, reviews etc.

**Research Assignments:**

* Researched modelling techniques for the dataset in project scope.
* Researched on cloud data pipeline steps and reviewed data engineer tutorials.

**Proof of Concept (POC):**

* Conducted POC on one modelling technique each.
* Conducted POC on data pipeline on Azure.
* Added weather data to existing data and performed POC on Azure pipeline using Azure Data factory.

**Implementation:**

* Continued working on cloud implementation with Azure.
* Used MySQL Server for database management.
* Ensured dataset stationarity for modelling using the Dickey-Fuller test.
* Demonstrated initial cloud implementation in Azure using data ingestion pipeline.
* Prepared a presentation on Azure implementation – key steps and methodologies.

**Testing and Further Implementations:**

* Tested cloud pipeline in parallel as a secondary task.
* Focused on testing simulation modelling for weather data.
* Implemented more resource utilization based on feedback.
* Listed benefits and drawbacks of models used for POC.

**Mid Semester Feedback:**

* Incorporated midterm feedback in further semester.
* Enhanced existing models with additional feature engineering techniques.
* Included more graphical representations and insights from simulations.
* Utilized Azure ML for final modelling process.

**Feature Engineering and Visualization:**

* Identified and applied relevant feature engineering techniques to improve model performance.
* Generated and incorporated additional visualizations to provide deeper insights.
* Integrated Azure ML for final model deployment and evaluation.

**Interim PPT Feedback:**

* Spread out key results into two slides instead of one.
* Refine insights and model slide to indicate different models tried and the reasons for choosing the final model for predicting bus delay time.

**Decisions Made:**

* Avoided running Azure ML workspace in the cloud due to costs.
* Assessed model readiness for consolidation.
* Final Checks on model for prediction on future values as well as by slicing current dataset to compare actuals with predictions.
* Final model to be published in cloud and compute instance to be stopped once model runs successfully to avoid additional costs.
* Report and presentation draft started in parallel with existing findings to summarize all insights till date.

|  |  |
| --- | --- |
| **Team Members** | **Project Closure Tasks Assignments** |
| Sankeerthana | Report Draft |
| Navya | PPT draft and consolidation. |
| Khyati | Finalise Azure ML model approach, report and ppt consolidation and documentation. |
| Udanie | Finalize model and code, ppt and report final version documentation. |
| Manisatya | Update interim ppt with latest findings. |

|  |  |  |
| --- | --- | --- |
| **Team Members** | **Overall Attendance** | **Key Contributions** |
| Khyati | 92% | Cloud data pipeline research and implementation, Azure ML readiness, Python Code troubleshooting on the go, Monte Carlo simulation testing, meeting minutes, interim ppt, key insights documentation. |
| Udanie | 70% | Modelling research and implementation of various techniques, model tuning, finalise the coding, interim report finalize |
| Sankeerthana | 70% | Initial research on cloud, meeting minutes, final report draft. |
| Navya | 70% | Final ppt draft, Weather data research and consolidation, cloud pipeline research and testing poc, meeting minutes, documentation of steps for cloud related work |
| Manisatya | 54% | Presentation drafting, Interim report drafting, unstructured dataset research |