**Team Meeting Minutes 1 – Monday 6th March 2023 6:00-7:00 pm (AEDT)**

**Location**: Remote via Discord

**Attendees**: Daniel, Mia, Kevin & Christian (Minute Taker)

**Items discussed:**

* Introductions (background/past data science experience/motivation to study MDS)
* Admin
  + Week 1 availabilities for 2nd team meet & meet with lecturer
    - Christian - unavailable Tuesday night
    - Mia - unavailable Thursday afternoon & Friday night
    - Kevin - preferred Friday night
    - Daniel - unavailable Wednesday night
  + Scheduled 2nd team meet - 9pm (AEDT) Thursday 9th March with all team members to attend
  + Scheduled meet with lecturer - Kevin enrolled our group into 7pm (AEDT) Friday 10th March with Wei Tian, consensus that this would be the best time to meet the lecturer as all team members will have completed Week 1 activities along with some progress made to Assessment 1 Part A: Group Project Plan  
      
    Zoom link to join: <https://unsw.zoom.us/j/81112831764>
  + Best form of ongoing team communication – private Discord server & Google Docs

**Action items:**

* Each team member to complete Week 1 activities
* Google Doc to be created for Assessment 1 Part A – Mia has created this with link to the Google Doc pinned in our discord chat & provided below[Assessment 1 Part A: Group Project Plan](https://docs.google.com/document/d/1kVH3z67ZRTm3L78suss2jHe4X1PhMs5H7Om4puxISI0/edit?usp=sharing)
* Google Doc to be created to record all Meeting Minutes – Christian has created this with link to the Google Doc pinned in our discord chat
* Team members to contribute to Assessment 1 Part A Google Doc

**Team Meeting Minutes 2 – Thursday 9th March 2023 9:00pm - 12:00am (AEDT)**

**Location**: Remote via Discord

**Attendees**: Daniel, Mia, Kevin & Christian (Minute Taker)

**Previous Action Items:**

* Confirm completion of previous Action Items
  + Each team member to complete Week 1 activities - Ongoing
  + Google Doc to be created for Assessment 1 Part A – Mia has created this with link to the Google Doc pinned in our discord chat & provided below[Assessment 1 Part A: Group Project Plan](https://docs.google.com/document/d/1kVH3z67ZRTm3L78suss2jHe4X1PhMs5H7Om4puxISI0/edit?usp=sharing)
  + Google Doc to be created to record all Meeting Minutes – Christian has created this with link to the Google Doc pinned in our discord chat
  + Team members to contribute to Assessment 1 Part A Google Doc - Ongoing

**Agenda/Discussion Items:**

* Friday-Sunday (10-12/3) & Week 2 availabilities
  + Christian - Free Friday afternoon, Saturday & Sunday
    - W2 free most evenings except Tuesday, free all Saturday & Sunday
  + Mia - Free Friday morning, Sunday evening, Monday morning
    - W2 free most days - Tuesday to Sunday
  + Daniel - Free Friday, Sunday early morning & late evening (AEDT)
    - W2
  + Kevin - Free Friday afternoon, Saturday afternoon, Sunday
    - W2
* Schedule next team meeting - Friday 6:15pm, Sunday 9pm (AEDT)
* Schedule/Book Week 2 meet with the lecturer - Friday 17 March 2023
  + https://unsw.zoom.us/j/81112831764
* Define roles and responsibilities for each member
  + Research Lead - Mia, Daniel
  + Data Analysts - Mia, Daniel, Kevin
  + Project Manager/Team Leader - Christian
* Establish teamwork software list - Python3 (Jupyter notebooks, all have experience), Tableau (Daniel, Kevin, Christian) & Power BI (Mia), Google Doc, Discord, MS Teams, MS Planner
* Project Management: GitHub (team access to repository) - Daniel <https://github.com/rgdk/UNSW-ZZSC9020-2023H2_project> , MS Teams & Planner
* Decide on list-making application(s) - MS Teams & Planner (chat, meetings, files, calendar, planner)
* Decide on file sharing method - MS Teams files, GitHub
* Decide on collaborative editing method - Google Docs (link in MS Team), Colab, Jupyter Notebooks share/merge through GitHub
* Decide on Group Meeting medium - Currently Discord, moving to MS Teams Conference
* Group Project Plan
  + Introduction and Motivation - All
  + Brief Literature Review - Daniel
  + Methods, Software and Data Description - Mia
  + Activities and Schedule - Christian (with assistance from all)

**Action Items:**

* For meet with Lecturer Wei Tian 7pm (AEDT) Friday 10th March 2023
* For Group Project Plan due 5pm (AEDT) Tuesday 14th March 2023
* Set-up MS Teams & Project Management related software (Christian)
  + Provide guide to assist MS Team/PM use
  + Gantt chart (in Teams)
* Data exploration (All)
* Preliminary research on topic, provide possible Research Questions (All)
* Finalise GitHub repository (Daniel & all)

**Team Meeting Minutes 3 – Friday 10th March 2023 7:00-8:00pm (AEDT)**

**Purpose:** Meeting with lecturer (Wei Tian) 7:30-8:00pm

**Location**: Remote via Discord (7:00-7:30pm), Zoom (https://unsw.zoom.us/j/81112831764)

**Attendees**: Daniel, Kevin & Christian (Minute Taker)

**Apologies:** Mia

**Previous Action Items:**

* For meet with Lecturer Wei Tian 7pm (AEDT) Friday 10th March 2023
* For Group Project Plan due 5pm (AEDT) Tuesday 14th March 2023
* Set-up MS Teams & Project Management related software (Christian)
  + Provide guide to assist MS Team/PM use
  + Gantt chart (in Teams)
* Data exploration (All)
* Preliminary research on topic, provide possible Research Questions (All)
* Finalise GitHub repository (Daniel & all)

**Agenda/Discussion Items:**

* Jupyter Notebooks + Word to produce the Group Project Report
* Wei confirmed we did not have to use RMarkdown, we need to justify why Python (Jupyter notebooks)
* Research Question - can be very simple but must be specific (can’t be too broad). RQ can change if need be
* Demand (all NSW), Temperature Bankstown only

**Action Items:**

* CB - try access/cloning Daniel’s private repository - able to access & clone repository but unable to upload/git push any files
* CB - finish setting up MS Teams & Planner Changed to ClickUp

**Team Meeting Minutes 4 – Sunday 12th March 2023 9:00-10:30pm (AEDT)**

**Location**: MS Teams video conference

**Attendees**: Daniel, Mia, Kevin & Christian (Minute Taker)

**Previous Action Items:**

**Agenda/Discussion Items:**

* Individual Skill Set Surveys
  + Mia completed
  + Kevin to complete tonight
  + Daniel to complete tonight
* Scheduled next meet for 7pm (AEDT) Monday 13th of March

**Action Items:**

* Daniel to work on the Introduction and Motivation and refine the literature review. DONE
* Mia to refine Methods, software and data description. DONE
* Christian to finish Activities and Schedule. DONE
* By 7pm Sydney time tomorrow

**Team Meeting Minutes 5 – Monday 13th March 2023 6:00-6:30pm (AEDT)**

**Location**: MS Teams video conference

**Attendees**: Daniel, Mia, Kevin & Christian (Minute Taker)

**Previous Action Items:**

* Daniel to work on the Introduction and Motivation and refine the literature review. Complete
* Mia to refine Methods, software and data description. Complete
* Christian to finish Activities and Schedule. Complete
* Mia, Kevin & Daniel to submit the skill set survey. Complete

**Agenda/Discussion Items:**

* Discussed editing Project Plan
* Edits were made, during conference, to the Literature Review and Methods sections

**Action Items:**

* Daniel to review/edit Project Plan to fit the designated 3-pages (locally) Complete
* Daniel to create a consistent style reference list Complete
* Final Project Plan Report to be viewed by all members Complete
* Final Project Plan Report to be submitted by Christian Complete
* Final Project Plan Report to be uploaded on GitHub (CB) Complete

**Team Meeting Minutes 6 – Friday 17th March 2023 6:00-6:40pm (AEDT)**

**Purpose:** Meeting with lecturer (Wei Tian) 6:00-7:00pm

**Location**: Zoom (https://unsw.zoom.us/j/81112831764)

**Attendees**: Daniel, Mia & Christian (Minute Taker)

**Apologies:** Kevin

**Previous Action Items:**

* Daniel to review/edit Project Plan to fit the designated 3-pages (locally) Complete
* Daniel to create a consistent style reference list Complete
* Final Project Plan Report to be viewed by all members Complete
* Final Project Plan Report to be submitted by Christian Complete
* Final Project Plan Report to be uploaded on GitHub (CB) Complete

**Agenda/Discussion Items:**

* Discussed Temp. Data, daily (mins & max)
* Wei believes we are on track
* Wei - our choice of model(s), justify by finding best use cases (seeing our project is limited to 6 weeks)
* Outside of meeting with Wei
  + Discussed setting up a recurring meeting, to meet at least twice a week plus meeting with lecturer on Friday.
  + Next Meet: 6pm (AEDT) Monday 20th of March (to confirm with Kevin too)
  + Booked Meeting with Wei 8-9pm (AEDT) Friday 24th of March

**Action Items:**

* Share GitHub repo with the all lecturers (CB)
* Team to finish up most of the Data Cleaning (by perhaps Mon/Tues 20/21 March)?
* Next week to have a good start on Data Exploration

**Team Meeting Minutes 7 – Monday 20th March 2023 6:00-6:40pm (AEDT)**

**Location**: MS Teams video conference

**Attendees**: Daniel, Kevin & Mia (Minute Taker)

**Apologies:** Christian

**Previous Action Items:**

* Share GitHub repo with the all lecturers - check with CB (CB completed)
* Team to finish up most of the Data Cleaning (by perhaps Mon/Tues 20/21 March)? - completed majority
* Next week to have a good start on Data Exploration - started time series plots

**Agenda/Discussion Items:**

* Temperature data
  + Average temperature not available for free
  + Use maximum temp - literature shows that max is often used - kevin to find
* Feature selection
  + Is rain worth including? Humidity/aircon, driers - increased machine use. Use humidity instead of rain? Keep unless we find humidity data
  + Exclude forecast demand until the end for comparison
  + Include solar installation output, not quantity
* Model selection - stacking (leveraging multiple models) recommended by literature review. Models showing decent results include:
  + SVM
  + Random forest
  + Ensemble - gradient boost, adaboost, etc.
  + Etc.
* Data exploration tasks
* Possible limitations
  + Utlising only Bankstown data for state demand - should include as a limitation in report, plus provide recommendations on how to improve analysis, e.g. utilise aggregated data from each postcode/LGA/region in NSW instead of just Bankstown
  + Rain is 100% humidity
* Next meeting

**Action Items:**

* Data exploration:
  + Outlier analysis
  + Correlation analysis
* Modelling
  + Implement standardisation (between 0 and 1) - all
  + Investigate stacking method - DK
  + Compare ensemble algorithms - MJ
  + Linear regression and random forest - KN
  + Decision tree - MJ/KN
  + Neural Networks/Deep Learning (e.g. LSTM) - DK (time consuming task)
* Next meeting: Thursday 23/03 - 7pm syd, 6pm bne, 10am israel

**Team Meeting Minutes 8 – Thursday 23rd March 2023 7:00-8:00pm (AEDT)**

**Location**: MS Teams video conference

**Attendees**: Daniel, Kevin, Christian & Mia (Minute Taker)

**Previous Action Items:**

**Agenda/Discussion Items:**

**Action Items:**

**Team Meeting Minutes 9 – Friday 24th March 2023 8:00 - 9:00pm (AEDT)**

**Location**: Zoom Meeting https://unsw.zoom.us/j/81112831764

**Attendees**: Daniel, Mia and Kevin (Minute Taker) with Lecturer Wei Tan

**Apologies:** Christian

**Questions and Answer with Lecturer**:

Q: Do we need to use one hot encoding to not losing marks?

A: No, dont need to. U can use whatever technique based on your data availability.

Q: Do we need to use technique like filter, wrapper or embedded in feature selection? Or just need to raise the reason behind the selection?

A: Just need to justify why you chose those features which should be based on previous paper.

Q: Do we need to do the exhausted literature review? Is that required?

A: It’s difficult to do. You should use the relevant paper to justify

It’s good to use particular model/method rather than apply the new one.

Doing on top of someone’s already done so more achievable.

Q: We do ensemble methods: random forest, linear regression, SVM, ANN … Is that ok?

A: Need to justify, provide cases for it, be able to explain and must be convincing rather than just simple analysis.

Q: Many models use ANN, and also other models then ended up with ANN.

A: Justification not only why you use particular method, but why or where this result comes from

Q: We forecast electricity demand for next 7 days?

A: Ok dont need to go with different time horizon

Should use one time horizon but good reasoning for that

Q: Which audience we should aim to?

A: You can choose your specific audience

Q: How many features are ok? We currently use 4-5?

A: You should be more than that

Q: Forecast demand for comparison’s sake at the end. Should we use Forecast demand in the model?

A: In the model, u can use Actual previous demand

If you use forecast demand, you would steal the old results of others.

Q: Looking at the modelling, 1 response variable and 8 variables in the vectors?

A: Maybe need to add in more

**Team Meeting Minutes 10 – Sunday 26th March 2023 8:30-9:45pm (AEDT)**

**Location**: MS Teams video conference

**Attendees**: Daniel, Kevin, Mia & Christian (Minute Taker)

**Previous Action Items:**

**Agenda/Discussion Items:**

* **Models** - LSTM Model & XGBoost apply SHAP to discuss/compare feature importance for day-ahead forecasting
* **Features to include** - day of the week (1-7), public holiday (0,1), wind speed, wind direction (if possible but not necessary)
* **Research Question** - “Explaining XGBoost versus LSTM models in day-ahead electricity demand forecasting” - discuss feature importance

**Action Items:**

* **Daniel** - Literature Review then help with LSTM
* **Mia** - XGBoost (adapt code from Gradient boost) then help Kevin with LSTM
* **Kevin** - LSTM (show some code research, build up model, prepare for writing about LSTM in the report)
* **Christian** - Put together Report/Marking Criteria, research SHAP method
  + [Group F Project Report (Draft)](https://docs.google.com/document/d/1ykYbR9JCMApprIjaExrGE5BZ61TvkJcY7WWskvnwTbo/edit?usp=sharing)
* Meeting with Gustavo Monday 5pm (4pm Bris) 27th March - <https://unsw.zoom.us/j/8978904335>
* Team Meeting Wednesday 7pm, 6pm (Bris), 10am (Israel) 29th March - CB to confirm with Daniel. Confirmed although Daniel will be in office.

**Team Meeting Minutes 11 – Monday 27th March 2023 5:00-6:00pm (AEDT)**

**Location**: Zoom video conference with Gustavo Batista - <https://unsw.zoom.us/j/8978904335>

**Attendees**: Kevin, Daniel, Mia & Christian (Minute Taker)

**Gustavo Points & Discussion Items**

* 10-12 variables/features is good - Gustavo reckons this is enough, especially because we have complimented them with outsourced data.
* Data scraping & how we got the data should be a focus in the report and in the presentation as well.
  + Gustavo would like to have access to our data for the purpose of providing this for future courses
* 3-5 years data is fine
* Recommend that we pick 1 or 2 time horizons. Gustavo believes we’ll have more to talk about for time horizons day-ahead or longer (7 days) instead of intra-day
* Weather data from bankstown only. Gustavo recommends just using the one temp location. Not much to gain from time spent getting other region temp.
* Gustavo really likes XGBoost, performs well and training the data doesn't take long, data-prep however may take a little longer as the data will need to be prepared in “slide windows” as XGBoost will want tables.
  + We’ll need to experiment with the “window size” for XGBoost
  + 50% slide windows are recommended
* LSTM - how many neurons, hidden layers etc. Training the model will be tedious. Will need a GPU.
* Gustavo really likes our selection of models. LSTM right model for this time of data, XGBoost a great model also with high accuracy (both are relative new models too)
* Make sure to compare with AEMO Model (as base Model). Forecast data provided. Will need to match the time forecast time horizons.
* Affirmation that our research question is great and novell (compared to other groups). Gustavo confirmed
* Gustavo recommended that we speak with Pierre who will like that our focus is on explainability - booked with Pierre for **Monday 3rd April 4:00pm (Syd/Bris) 8am (Israel)** <https://unsw.zoom.us/j/86520081658>
  + Would be a good idea going into this meet with questions around SHAP/explainability hurdles etc

**Team Meeting Minutes 12 – Wednesday 29th March 2023 7:00-7:40pm (AEDT)**

**Location**: MS Teams video conference

**Attendees**: Mia & Christian (Minute Taker)

**Apologies:** Daniel & Kevin

**Previous Action Items:**

* **Daniel** - Literature Review then help with LSTM
* **Mia** - XGBoost (adapt code from Gradient boost) then help Kevin with LSTM
* **Kevin** - LSTM (show some code research, build up model, prepare for writing about LSTM in the report)
* **Christian** - Put together Report/Marking Criteria, research SHAP method
  + [Group F Project Report (Draft)](https://docs.google.com/document/d/1ykYbR9JCMApprIjaExrGE5BZ61TvkJcY7WWskvnwTbo/edit?usp=sharing)

**Agenda/Discussion Items:**

* Easter Holidays plans (Friday 7th to Monday 10th of April).

Note, this is the last weekend before the Report is Due (10pm AEST Saturday 15th April)

* + Mia, around
  + Christian, around
  + Kevin?
  + Daniel?
* Next Team Meeting for Saturday 1st of April? Afternoon (AUS) / Morning (Israel)?
* Reminder, meeting with Pierre is **Monday 3rd April 4:00pm (Syd/Bris) 8am (Israel)** <https://unsw.zoom.us/j/86520081658>

**Action Items:**

* **Mia**
  + Start drafting Materials & Methods (Software, Description of the Data, Pre-processing Steps, Data Cleaning) leverage Project Plan
  + Start drafting Modelling Methods (XGBoost)
    - We will need a deep understanding of the model to provide a detailed description of the model, explain any assumptions, discuss math/statistic/algorithm used
    - Model structure and tuning
    - Hyperparameters
    - Model assessment / diagnostic
  + EDA - highlight main observations you have uncovered
* **Kevin** 
  + Start drafting Modelling Methods (LSTM)
    - We will need a deep understanding of the model to provide a detailed description of the model, explain any assumptions, discuss math/statistic/algorithm used
    - Model structure and tuning
    - Hyperparameters
    - Model assessment / diagnostic
  + EDA - highlight main observations you have uncovered
* **Daniel** 
  + Finalise literature review, also include literature specifically around XGBoost & LSTMs and what they are? Literature around SHAP?
  + Assist with Model tuning, assessment / diagnostic for both XGBoost & LSTM
* **Christian**
  + Develop a deep understanding of SHAP
    - Source multiple papers around SHAP (could provide this for literature review)
    - SHAP will form an integral part in Analysis and Results section and/or Discussion section
      * base models analysis/results > SHAP analysis of base models > improve models based off SHAP findings > improved models analysis/results **OR**
      * tuned & performing models > SHAP analysis to build understanding of inner workings of models

**Team Meeting Minutes 13 – Sunday 2nd April 2023 7:30pm (AEST)**

**Location**: MS Teams video conference

**Attendees**: Daniel, Kevin, Mia & Christian (Minute Taker)

**Apologies:**

**Previous Action Items:**

* **Mia**
  + Start drafting Materials & Methods (Software, Description of the Data, Pre-processing Steps, Data Cleaning) leverage Project Plan
  + Start drafting Modelling Methods (XGBoost)
    - We will need a deep understanding of the model to provide a detailed description of the model, explain any assumptions, discuss math/statistic/algorithm used
    - Model structure and tuning
    - Hyperparameters
    - Model assessment / diagnostic
  + EDA - highlight main observations you have uncovered
* **Kevin** 
  + Start drafting Modelling Methods (LSTM)
    - We will need a deep understanding of the model to provide a detailed description of the model, explain any assumptions, discuss math/statistic/algorithm used
    - Model structure and tuning
    - Hyperparameters
    - Model assessment / diagnostic
  + EDA - highlight main observations you have uncovered
* **Daniel** 
  + Finalise literature review, also include literature specifically around XGBoost & LSTMs and what they are? Literature around SHAP?
  + Assist with Model tuning, assessment / diagnostic for both XGBoost & LSTM
* **Christian**
  + Develop a deep understanding of SHAP
    - Source multiple papers around SHAP (could provide this for literature review)
    - SHAP will form an integral part in Analysis and Results section and/or Discussion section
      * base models analysis/results > SHAP analysis of base models > improve models based off SHAP findings > improved models analysis/results **OR**
      * tuned & performing models > SHAP analysis to build understanding of inner workings of models

**Agenda/Discussion Items:**

* Easter Holidays plans (Friday 7th to Monday 10th of April).

Note, this is the last weekend before the Report is Due (10pm AEST Saturday 15th April)

* + Mia, around (except Friday)
  + Christian, around
  + Kevin, around (except Sunday)
  + Daniel, around (except Saturday)
* Reminder, meeting with Pierre is **Monday 3rd April 4:00pm (Syd/Bris) 9am (Israel)** <https://unsw.zoom.us/j/86520081658>
* Data: 4 years training 1 Aug 2017 >, 1 year testing (Aug 21-22)
* Individual Reflective Portfolio (5%) **Due 5pm(Aus)/10am(Israel) Tuesday 4th April**

**Action items:**

* Kevin:
  + Update model & upload code (GitHub), having issues with running the model (GridSearch). Daniel to review & assist.
  + Add notes to contain trial & error testing for number of layers, parameter tuning etc (link to papers if possible) [Try to be as detailed as possible]
  + Surface level explanation on LSTMs (algorithm/stats/math)
* Mia:
  + Tune XGBoost (look-back etc),
  + Surface level explanation on LSTMs (algorithm/stats/math)
  + If you have time, start on:
    - 3.3 Pre-processing Steps
    - 3.4 Data Cleaning
* Daniel:
  + Review LSTM code, assist Kevin with LSTM (Gridsearch) model
  + Review XGBoost code & assist Mia
  + Start on EDA, include comments around extremes etc
* Christian:
  + Finish SHAP review
  + Apply SHAP on LSTM (once model is complete)
    - Start providing dot-points/tables/graphs for 5 Analysis & Results
  + Apply SHAP on XGBoost (once model is complete)
    - Start providing dot-points/tables/graph for 5 Analysis & Results

**Team Meeting Minutes 14 – Monday 10th April 2023 8:20-9:20pm (AEST)**

**Location**: MS Teams video conference

**Attendees**: Daniel, Kevin & Christian (Minute Taker)

**Apologies:** Mia

**Availability (10th - 15th April) - Report Due 10pm (AEST) Saturday 15th April**

* Kevin - available most nights
* Daniel - Tues night & Weds day unavailable (Israel time)
* Christian - available most nights with full availability Weds & Fri
* Mia - available Tues, Weds, Thurs, Sat. Unavailable Friday.

**Action items:**

* Kevin:
  + Finish 3.6.2 LSTM under Modelling Methods- DONE
  + Finish 5.3 LSTM under Analysis & Results for V1,V2 & V3- DONE
  + If you have time, jot down a few points around feature importance outcome for LSTM (Why TOTALDEMAND was no.1? etc) under Discussion
* Mia:
  + Finalise XGBoost related sections (Modeling Methods, Analysis & Results, maybe Discussion?)
* Daniel:
  + Finish EDA Provide commentary around Mia’s EDA graphs/plots etc - DONE
  + Finish Introduction - DONE
  + Finish Lit Review to include XAI methods + Approx. Model for feature importance - DONE
  + If you have time, jot down a few points around feature importance outcome for LSTM (Why TOTALDEMAND was no.1? etc) under Discussion = DONE
  + Also re-coded v3 of the LSTM Bayesian Optimisation notebook - DONE
* Christian:
  + Finish SHAP under Modelling Methods (3.6.3)
  + Finish SHAP Analysis for XGBoost 5.2.2 under Analysis & results
  + Once Kevin/Daniel have completed LSTM V3, I’ll finish Feature Importance 5.3.2 under Analysis & Results
  + Discussion - with backing from literature, provide content around the meaning for each model’s feature importance outcomes.
  + Start planning the Structure of the Presentation
    - **Due 5pm (AEST) Tuesday 18th April**