PROJECT PLAN BY GROUP B

FORECASTING THE TRANSITION: ANALYSING THE FEASIBILITY FOR RENEWABLE ENERGY ADOPTION IN NSW

Kiara Zentrich (z5382804)

Mathew Fraser(z5433663)

Santosh Ban (z5441817)

Karunya Sankar Kumar (z5339860)

March 2024

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS OF THE CAPSTONE COURSE ZZSC9020

Abstract

Electricity serves as a cornerstone of modern civilisation, powering industries, homes, and critical infrastructure. However, the reliance on non-renewable sources of electricity generation poses significant environmental challenges, exacerbating climate change and endangering ecosystems. In response, the transition to renewable energy sources has emerged as a pressing priority worldwide. This project addresses the transition to a fully renewable energy-based electricity supply in New South Wales (NSW), Australia. The Ecological Defenders Office (EDO), a prominent environmental advocacy firm in NSW, seeks evidence-based forecasts to assess the feasibility of achieving this transition within the next decade. Specifically, the project aims to determine when renewable energy generation can meet forecasted electricity demand in the region. The significance of this endeavour cannot be overstated. Transitioning to renewable energy not only mitigates environmental harm but also fosters energy independence and resilience. The methodology for this project involves several key steps. Firstly, historical electricity demand data will undergo preprocessing to clean, detect outliers, and impute missing values. External factors such as weather data will be incorporated to enhance forecasting accuracy. Exploratory data analysis will identify relevant features, trends, and seasonal patterns. Supervised machine learning models, particularly regression and classification, will be employed for forecasting, considering the non-linear nature of the data. The project will utilise Python programming language and libraries such as scikit-learn and matplotlib for implementation. Google Colab will provide a collaborative development environment, ensuring reproducibility and transparency. Lastly, the data used for analysis will comprise historical electricity demand data, weather data, and additional external factors. This data, while comprehensive, may require normalisation and transformation to facilitate accurate modelling. By employing a combination of statistical and machine learning techniques, this project aims to deliver actionable insights into the feasibility of transitioning to renewable energy in NSW. The findings will inform advocacy efforts and policy decisions, driving progress towards a sustainable energy future. By providing the EDO with accurate forecasts, this project empowers them to advocate for policies and investments that accelerate the transition to renewable energy in NSW.

Contents

1 Introduction and Motivation

Electricity is a fundamental form of energy resulting from the existence of charged particles (such as electrons and protons) and their interactions. It is a versatile and widely used source of power that has transformed industries, communication, and daily life, contributing to global progress, and shaping the way we live and work. Produced through various means, including fossil fuels, nuclear reactions, and renewable sources, electricity is transmitted through power grids to homes, businesses, and industries.

However, it is now globally recognised that non-renewable sources of electricity generation, particularly fossil fuels (namely coal, natural gas, and oil) have significant and sustained deleterious effects on the environment (Chowdhury & Oo, 2012)

Continued use of these sources would further exacerbate climate change, jeopardising ecosystems, threatening the survival of species, and undermining the health and well-being of present and future generations (McCarty, 2001). As a result, renewable sources, such as wind, solar and hydro, have emerged as clean alternatives to generate electricity whilst reducing dependency on fossil fuels and greenhouse gas emissions.

Transforming the electricity network into a 100% renewable energy scenario is a crucial, but challenging priority in New South Wales (NSW). Achieving this target as soon as possible will result in a greater reduction of potential climate impacts within the region. However, a key concern is whether renewable sources can reliably provide enough electricity to meet NSW demand. The aim of this project is to provide a key environmental advocacy firm in NSW, The Ecological Defenders Office, with evidence-based forecasts to answer the following question: Is the transition to a fully renewable energy-based electricity supply achievable within NSW in the next 10 years, and at what point will renewable energy generation meet the forecasted demand for electricity in the region?

The output from this analysis and modelling will empower the Ecological Defenders Office (EDO) to make informed advocacy and funding decisions. NSW currently has close to 13,500 megawatts of renewable energy generation capacity, which is 53% of total generation capacity in the state (Government, 2023). The EDO recognise this as strong progress, however achieving a fully clean energy state remains a way off, with the current capacity representing just over halfway toward that goal. The EDO has commissioned this work to understand the earliest point in time at which achieving a 100% renewable electricity target is feasible. By determining this timeline, they aim to devise effective lobbying strategies for governments and electricity providers. If forecasts indicate that renewables can meet demand earlier in the 10-year period, the organisation can prioritize shorter-term, high-intensity campaigns. However, if the forecasts suggest feasibility closer to the 10-year mark, longer-term advocacy requiring extensive planning and organisation will be necessary (Development, n.d.)

This project aims to assess the viability of transitioning to a fully renewable electricity supply in NSW within a decade. Feasibility will be determined based on the analysis indicating whether electricity supply from renewable sources can meet predicted demand levels. The project methodology involves the use of data science techniques to (1) forecast electricity demand in NSW for the next decade and then (2) forecast the supply of electricity solely from renewable sources in NSW over the same period. These forecasts can be generated using two multivariate regressions with the target variables as electricity demand and renewable electricity supply respectively. Finally, the project will then integrate the results from these models to determine the point at which supply can effectively match demand.

2 Brief Literature Review

Our research attempts to forecast electricity demand for NSW in the next decade. Therefore, resources from the following papers could provide insights on what type of models can be used for forecasting electricity demand.

In the study conducted by CSIRO, they have used smart metering and weather data to develop machine learning models that provide hour-ahead forecasts of aggregate demand for regions of geospatially proximate houses. These approaches excel in capturing non-linear relationships, handling large datasets, and adapting to changing conditions. The Self-tuning Support Vector Regression with RBF kernels can achieve an R^2 value of 0.95 and an Average MAE value as low as 0.02. There is a growing trend toward employing machine learning techniques, including neural networks, support vector machines, and decision trees, to enhance forecasting accuracy (Nguyen & Berry, 2019).

Currently, the Australian Energy Market Operator (AEMO) generate models using a selection of machine learning algorithms. These algorithms include LASSO, Gradient Boosting, Decision Trees and Random Forests. In developing the models, a series of diagnostic checks are performed and if the models fail these checks, the algorithm is adjusted iteratively. These diagnostic checks include k-folds out-of-sample cross validation, checking for under or overfitting by examining disparities between in-sample and out-of-sample accuracies, verifying appropriateness of variables by inspecting relationship between variables and residuals, observing residuals at extremes of demand to ensure assumptions for maximum and minimum demand are valid and no bias exists in these regions. Finally, predictions are compared to actual data and against detrended historical data (Australian Energy Market Operator, 2023).

The following work is relevant to our research as we are looking to forecast the supply of electricity from renewable sources in NSW. A research article from the Heliyon furnishes a forecasting model for both energy consumption and generation based on real data captured from a P2P grid system in the state of Western Australia. The methodology involves a comprehensive review of existing literature on short-term forecasting techniques for renewable energy consumption and generation. Various forecasting methods, including statistical models, machine learning algorithms, and hybrid approaches, are evaluated based on their performance metrics, computational efficiency, and applicability to regional factors such as weather patterns and energy demand profiles.

The results of the case study highlight the performance of various forecasting techniques in predicting short-term renewable energy consumption and generation in Western Australia. Statistical models such as autoregressive integrated moving average (ARIMA) and exponential smoothing methods are compared with machine learning algorithms like artificial neural networks (ANNs) and support vector machines (SVMs). The discussion emphasizes the strengths and limitations of each approach and proposes recommendations for improving forecasting accuracy and robustness (Abu-Salih, et al., 2022).

Ben Elliston's study conducts simulations to explore scenarios where 100% of electricity in the Australian National Electricity Market comes from renewable sources. The outcome of this paper is relevant for our work as we are trying to predict the year in which NSW can achieve 100% renewably sourced electricity. The research examines various aspects such as the feasibility, challenges, and implications of achieving such a transition. It delves into factors like renewable energy variability, grid stability, required infrastructure changes, and potential policy interventions necessary to facilitate the shift towards a fully renewable electricity system. The study provides insights into the technical, economic, and environmental considerations associated with transitioning to 100% renewable electricity in Australia's energy market (Elliston, Diesendorf, & MacGill, 2012).

Figure 1 below shows all the Artificial Intelligence types and algorithms available (Benti, Chaka, & Semie, 2023).

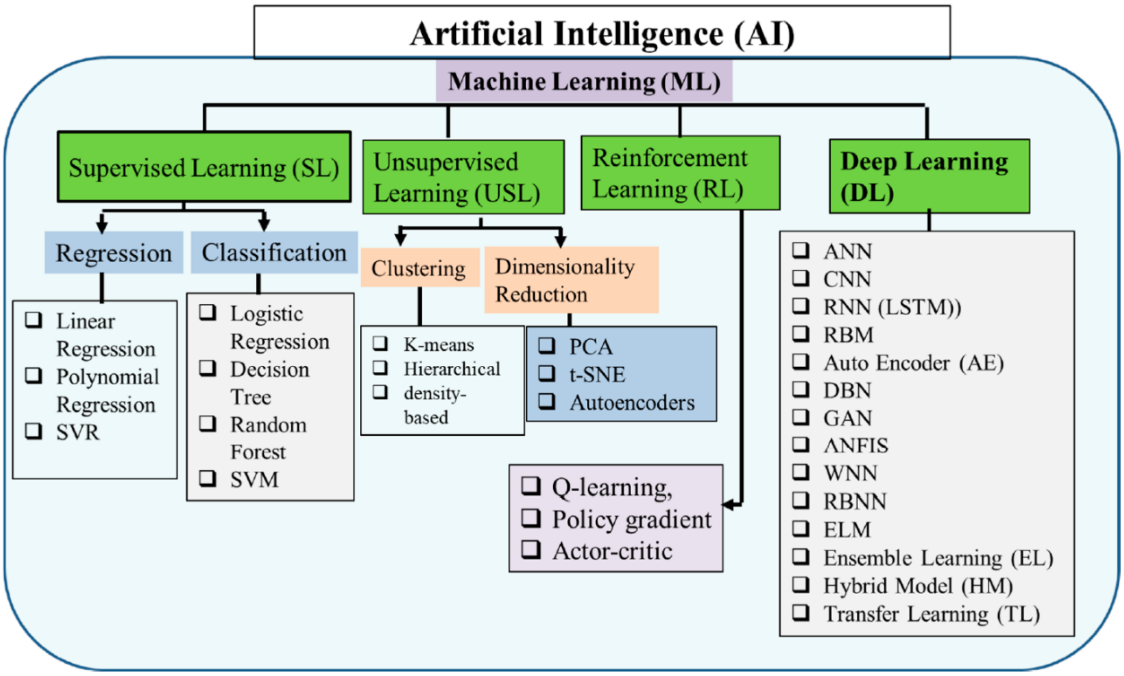


Figure 1 Machine Learning Model Options

Linear regression serves as a solid baseline model choice, while random forest, XGBoost, and support vector machines (SVMs) excel in managing non-linear relationships and intricate data structures. While classical machine learning (ML) models can be employed for forecasting, they may fall short compared to specialized time-series forecasting techniques like autoregressive models, moving averages, and recurrent neural networks (RNNs). The primary reason for this discrepancy lies in classical ML models' assumption of independent and uniformly distributed data points, a feature often absent in time series data. Time series data is characterized by temporal dependencies, where values at one time point are influenced by those at preceding points. The independence assumption challenges classical ML models' ability to discern underlying patterns and trends. Additionally, classical ML models are ill-equipped to handle time-varying features or unevenly spaced time series data, common occurrences in forecasting tasks. For instance, classical ML models may struggle to capture long-term seasonality trends within the data (Benti, Chaka, & Semie, 2023).

3 Methods, Software and Data Description

The following sequence of steps explain the methodology for this research.

Data Pre-processing: The historical electricity demand data provided for this project will be pre-processed to perform data cleaning, outlier detection and missing value imputation. Any relevant external factors such as weather data (provided by the course), public holidays and weekend calendar data will be incorporated into the data set.

Exploratory Data Analysis and Feature engineering: Important features such as trends and seasonal indicators will be identified. If required, categorical variables will be transformed into proper formats and continuous features might be normalised. The model selection (e.g. type of regression used) will be done depending on whether the variables are linear or non-linear.

Model Selection: Various models can be employed for forecasting electricity demand, depending on the level of detail required, the availability of data, and the forecasting horizon. To answer the research question, a thorough literature review will be done to understand various methods to forecast future electricity demand. From the brief literature review conducted for the Project plan it is understood that classical, supervised, and deep learning-based models can be used for forecasting (Korstanje, 2023).

For this work, the initial decision has been made to use **Supervised Machine Learning models such as Regression and Classification**. While literature review suggests, Supervised ML methods are not the first choice when it comes to time series data and prediction, they can still be applied effectively, especially when dealing with certain types of time series data or when there's a need to incorporate various external factors. A baseline model maybe built using linear regression and more models will be built to compare the performance.

Training and Validation: Splitting the dataset into training, validation and test sets is necessary. Training the selected models on training data and tuning hyperparameters using the validation set. Then the model performance can be monitored on the validation set to avoid overfitting.

Ensemble Modelling: If time permits, we will construct ensemble models by combining predictions from multiple individual models using techniques such as stacking.

Evaluation metrics: Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) will be calculated to assess the accuracy of the forecasting models. The metrics from different models will then be compared to pick the best performing model.

Analysis and discussion of results: Comparing the Forecasted demand value with the actual observations will help demonstrate the effectiveness of the model in predicting demand patterns.

Conclusion and Future Work: This part will summarise the key findings of the proposed methodology. We will also provide recommendation for the client on how to use the research to further their ambitions. This will also be a chance to propose potential avenues for future research to address emerging challenges in renewables replacing fossil fuels in electricity generation.

Software:

This section outlines the software tools, libraries, and platforms that will be used for implementing the electricity demand forecasting methodology. It emphasizes the use of open-source software, reproducibility, and accessibility to promote transparency and collaboration.

Development Environment: The paper will be implemented using Python programming language, leveraging its rich ecosystem of libraries for data analysis, machine learning and timeseries forecasting.

Libraries: The following libraries will be implemented in the code developed for this project

- Pandas and NumPy will be used to manipulate the data and perform any necessary operations to examine or transform the data.
- Scikit-Learn will be used to generate the regression models that will be employed for the forecasting methods.
- The Matplotlib and Seaborn libraries will be used to create any visualisations needed for data analysis or for the explanation and discussion of results.

Integration and Workflow: A Jupyter notebook hosted in Google Colab will be used to develop and execute the code. Colab has been chosen as it is cloud based which allows for ease of access to all team members, as well as providing GPU and TPU resources which will allow complex models to be run more efficiently than on any of the team's local resources. Colab will also enable easy version control as it can be connected to the GitHub repository being used by the team.

Data Description

At this stage of the project, five datasets have been selected for use in the analysis and modelling involved in this forecasting electricity demand research. The details and relevance of each are below:

Dataset: **Total Electricity Demand (NSW)**	
Source	Market Management System database, National Energy Market
Format	CSV
Storage	5.8MB
Variables	DATETIME: Date and time interval of each observation in the format (dd/mm/yyyy hh:mm). In 5 minute increments. TOTALDEMAND: Total demand (MW) REGIONID: Region Identifier (i.e. NSW1)
Messiness	NA: 0 Duplicates: 0 Inconsistencies: • Format of DATETIME is object. In cleaning, should change to a consistent datetime data type for easier manipulation.
Size	196513 rows x 3 columns
Relevance	This data shows historical electricity demand in NSW from 2010-01-01 to 2021-03-18. Historical demand is a powerful feature, which can be used as a predictor variable in electricity forecasting models.

Dataset: **Air Temperature (NSW)**	
Source	Australian Data Archive for Meteorology
Format	CSV
Storage	6.9MB
Variables	DATETIME: Date time interval of each observation (dd/mm/yyyy hh:mm) TEMPERATURE: Air temperature (°C) LOCATION: Location of a weather station (i.e. Bankstown weather station)
Messiness	NA: 0 Duplicates: 13 duplicated rows Inconsistencies: • Format of DATETIME is object. In cleaning, should change to a consistent datetime data type for easier manipulation. • The smallest gap between temperature observations is 1 minute and the largest gap is 3 days, 18hrs and 30mins.
Size	220326 rows x 3 columns
Relevance	Air temperature data covering 2010-01-01 to 2021-03-18 can also be considered as a potential predictor variable in forecasting. This

	data will be examined more deeply in the exploratory data analysis stage to identify if there is indeed a correlation between temperature and electricity usage. If this is proven, the data will form part of the modelling work undertaken in this project.

Dataset: **Population and Projected Population Totals (persons), 1971-2061**	
Source	2022 NSW Common Planning Assumption Projections, NSW Government
Format	XLSX
Storage	1.4MB
Variables	YEAR: The calendar year for each population projection (yyyy) POPULATION: Number of total persons living in NSW in a given year
Messiness	NA: 0 Duplicates: 0 Inconsistencies: 0
Size	91 rows x 2 columns
Relevance	This data covers NSW population from 1971 to 2023, and then projected population to 2061. This data will be examined more deeply in the exploratory data analysis stage to identify if there is a relationship between population and electricity demand. If this is proven, the data will form part of the modelling work undertaken in this project.

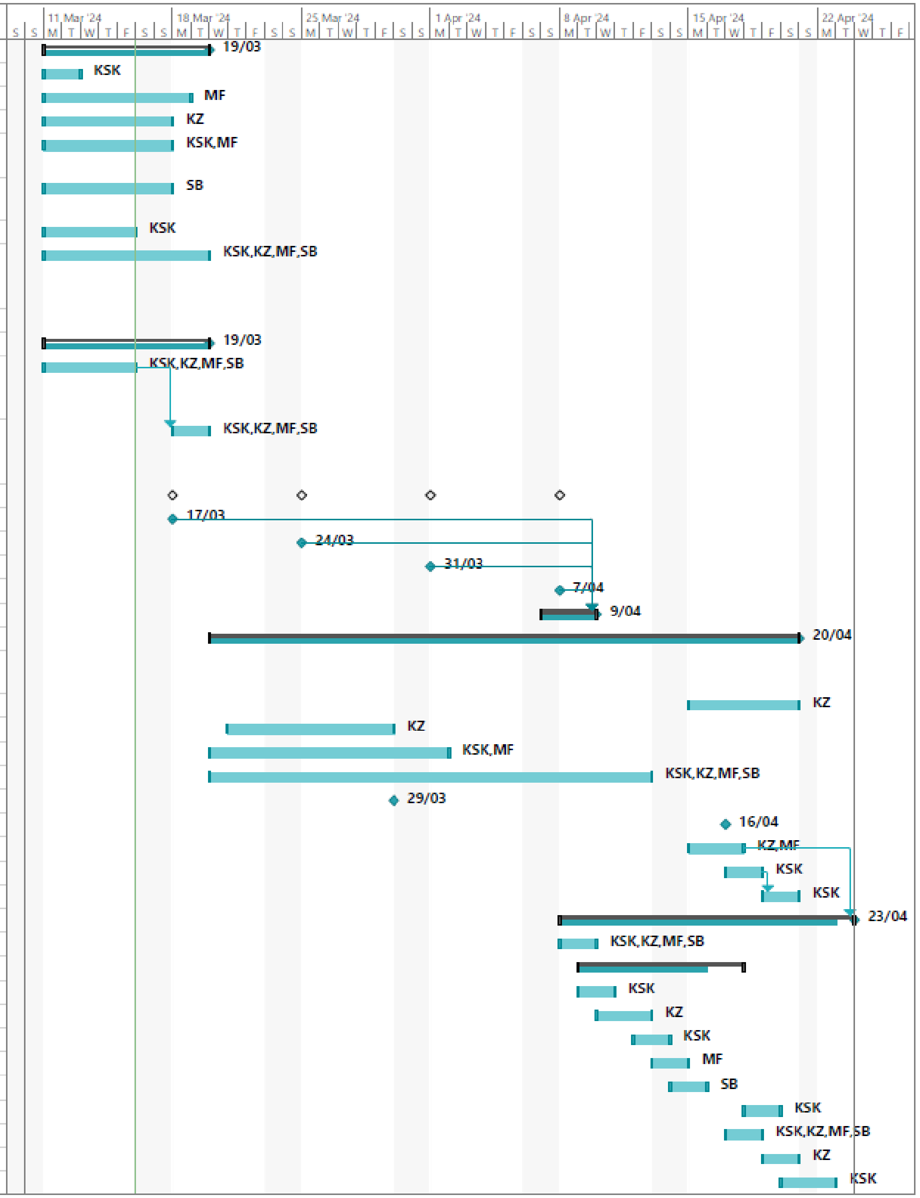
Dataset: **Forecasted Demand (NSW)**	
Source	Market Management System database, National Energy Market
Format	CSV
Storage	739.6MB
Variables	DATETIME: Date time interval of each observation (dd/mm/yyyy hh:mm). In half-hourly increments. FORECASTDEMAND: Forecast demand (MW) REGIONID: Region Identifier (i.e. NSW1) PREDISTPATCHSEQNO: Unique identifier of predispatch run (YYYYMMDDPP) PERIODID: Period count, starting from 1 for each predispatch run. LASTCHANGE: Date time interval of each update of the observation (dd/mm/yyyy hh:mm)
Messiness	NA: 0 Duplicates: 284 duplicated rows

	Inconsistencies: • Format of DATETIME and LASTCHANGED are an object. In cleaning, should change to a consistent datetime data type for easier manipulation.
Size	10906019 rows x 6 columns
Relevance	This data contains forecasted electricity demand in NSW from 2010-01-01 to 2021-03-18. This data can serve as an effective validation dataset, enabling the team to assess the accuracy of the predictive forecast models developed.

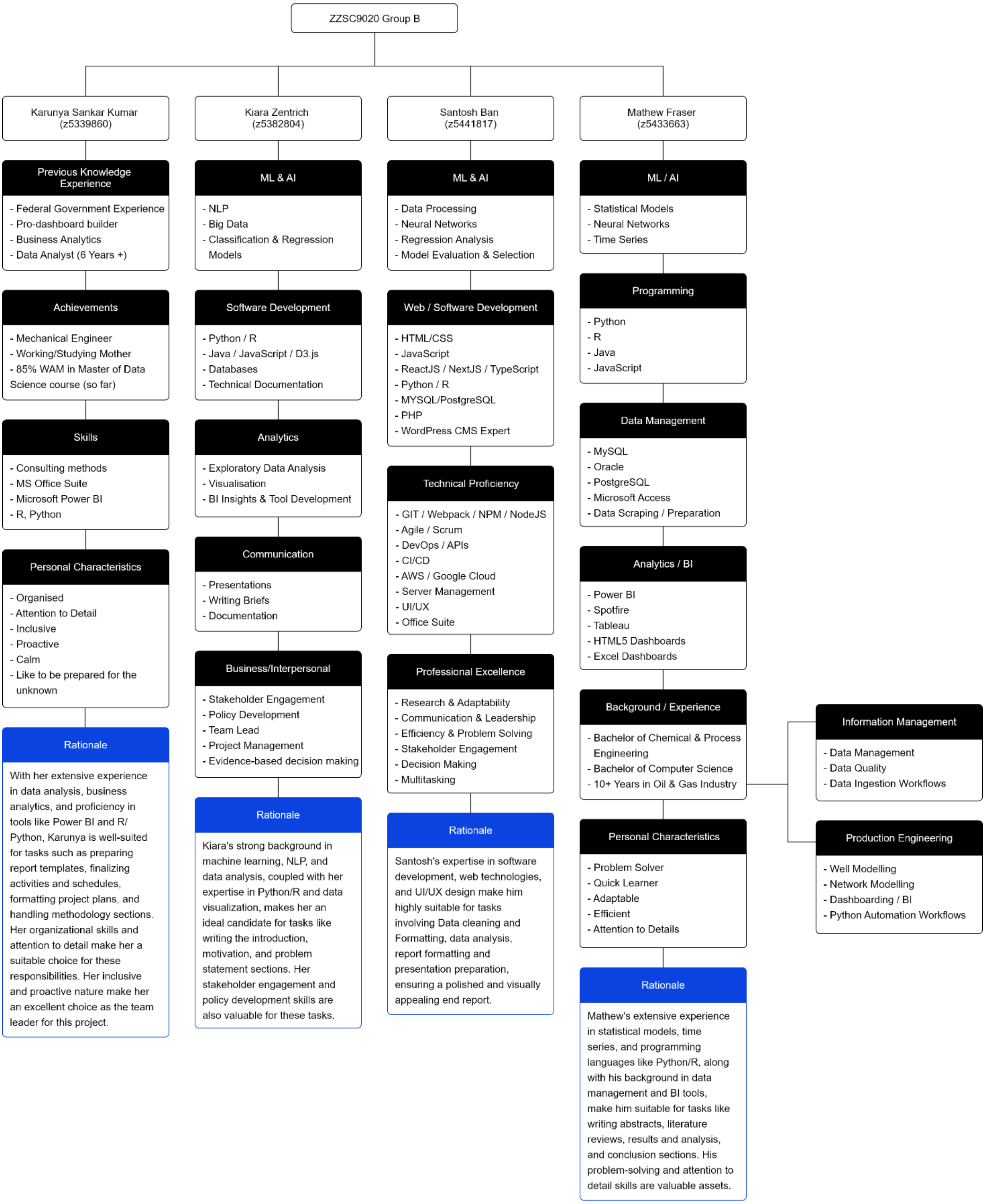
Dataset: **Electricity generation in New South Wales, by fuel type, physical units, financial year**	
Source	Department of Climate Change, Energy, the Environment and Water, Australian Energy Statistics, Australian Government
Format	XLSX
Storage	150KMB
Variables	YEAR: The financial year corresponding to the amount of electricity generated in that time period (yyyy-yy) FUEL SOURCE: The type of energy source (ie. Types of renewable and non-renewable fuels) ELECTRICITY GENERATED: The amount of electricity generated by fuel type (GWh)
Messiness	NA: 0 Duplicates: 0 Inconsistencies: • Years in this dataset are financial years, whereas other data to be used in this project is generally in the format of calendar year.
Size	13 rows x 14 columns
Relevance	This data details the amount of electricity generated by each fuel source from 2008-09 to 2021-22 in NSW. This data is key to the analysis and modelling that will be conducted to assess the feasibility of a 100% renewable electricity scenario within 10 years. This historical data can serve as a predictor variable in the modelling.

4 Activities and Schedule

ID	Task Name	Duration	Start	Finish	Predece	Resource Names
1	**1A. Group Project Plan**	**7 days**	**Mon 11/03/24**	**Tue 19/03/24**		**KSK,KZ,MF,SB**
2	Prepare Report Template	2 days	Mon 11/03/24	Tue 12/03/24		KSK
3	Write Abstract	6 days	Mon 11/03/24	Mon 18/03/24		MF
4	Write Introduction and Motivation	6 days	Mon 11/03/24	Sun 17/03/24		KZ
5	Write Literature Review	6 days	Mon 11/03/24	Sun 17/03/24		KSK,MF
6	Write Methods, Software and Data Description	6 days	Mon 11/03/24	Sun 17/03/24		SB
7	Finalise Activities and Schedule	5 days	Mon 11/03/24	Fri 15/03/24		KSK
8	Continuously update references	7 days	Mon 11/03/24	Tue 19/03/24		KSK,KZ,MF,SB
9	Format Project Plan					KSK
10	**1B. Individual Implementation Check**	**7 days?**	**Mon 11/03/24**	**Tue 19/03/24**		**KSK,KZ,MF,SB**
11	Discuss tools we are going to use as a team	5 days	Mon 11/03/24	Fri 15/03/24		KSK,KZ,MF,SB
12	Complete the online questionnaire (individually)	2 days	Mon 18/03/24	Tue 19/03/24	11	KSK,KZ,MF,SB
13	**2. Weekly Reflection**	**17 days**	**Sun 17/03/24**	**Tue 9/04/24**		**KSK,KZ,MF,SB**
14	2. Weekly Reflection 1	1 day	Sun 17/03/24	Sun 17/03/24		KSK,KZ,MF,SB
15	2. Weekly Reflection 2	1 day	Sun 24/03/24	Sun 24/03/24		KSK,KZ,MF,SB
16	2. Weekly Reflection 3	1 day	Sun 31/03/24	Sun 31/03/24		KSK,KZ,MF,SB
17	2. Weekly Reflection 4	1 day	Sun 7/04/24	Sun 7/04/24		KSK,KZ,MF,SB
18	2. Individual Reflection Portfolio	3 days	Sun 7/04/24	Tue 9/04/24	14,15,1	KSK,KZ,MF,SB
20	**3. Group Project Report**	**24 days?**	**Wed 20/03/24**	**Sat 20/04/24**		**KSK,KZ,MF,SB**
21	Plagiarism Statement and Acknowledgement					KSK
22	Abstract	6 days	Mon 15/04/24	Sat 20/04/24		KZ
23	Introduction	7 days	Thu 21/03/24	Fri 29/03/24		KZ
24	Literature Review	9 days	Wed 20/03/24	Mon 1/04/24		KSK,MF
25	Materials and Methods	18 days	Wed 20/03/24	Fri 12/04/24		KSK,KZ,MF,SB
26	Exploratory Data Analysis	8 days	Wed 20/03/24	Fri 29/03/24		SB
27	Analysis of results	20 days	Wed 20/03/24	Tue 16/04/24		KSK,KZ,MF,SB
28	Discussion, Conclusion and further iss	3 days	Mon 15/04/24	Wed 17/04/24		KZ,MF
29	Format Report	2 days	Wed 17/04/24	Thu 18/04/24		KSK
30	Submit Report	2 days	Fri 19/04/24	Sat 20/04/24	29	KSK
31	**4. Group Video Presentation**	**13 days?**	**Mon 8/04/24**	**Tue 23/04/24**	**28**	**KSK,KZ,MF,SB**
32	Decide Video recording Tool	2 days	Mon 8/04/24	Tue 9/04/24		KSK,KZ,MF,SB
33	**Prepare Presentation Slides**	**7 days**	**Tue 9/04/24**	**Wed 17/04/24**		**KSK,KZ,MF,SB**
34	Prepare Presentation Template	2 days	Tue 9/04/24	Wed 10/04/24		KSK
35	Problem Statement	3 days	Wed 10/04/24	Fri 12/04/24		KZ
36	Methodology	2 days	Fri 12/04/24	Sat 13/04/24		KSK
37	Results and Analysis	2 days	Sat 13/04/24	Sun 14/04/24		MF
38	Conclusions	2 days	Sun 14/04/24	Mon 15/04/24		SB
39	Coordinate recording times	2 days	Thu 18/04/24	Fri 19/04/24		KSK
40	Record video	2 days	Wed 17/04/24	Thu 18/04/24		KSK,KZ,MF,SB
41	Edit video together	2 days	Fri 19/04/24	Sat 20/04/24		KZ
42	Submit Video	2 days	Sat 20/04/24	Mon 22/04/24		KSK



5 Team Skill Mapping



6 References

Abu-Salih, B. et al., 2022. Short-term renewable energy consumption and generation forecasting: A case study of Western Australia. *Heliyon 8,* 8(3).

Ahmad, T., Zhang, H. & Yan, B., 2020. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustainable Cities and Society,* Volume 55, p. 102052.

Australian Energy Market Operator, 2023. *Forecasting Approach - Electricity Demand Forecasting Methodology,* s.l.: s.n.

Benti, N. E., Chaka, M. D. & Semie, A. G., 2023. Forecasting Renewable Energy Generation withMachine Learning. *Sustainability,* Volume 15.

Chowdhury, S. H. & Oo, A. M. T., 2012. Study on electrical energy and prospective electricity generation from renewable sources in Australia. *Renewable and Sustainable Energy Reviews,* 16(9), pp. 6879-6887.

Development, C. f. C. H. a., n.d. *Chapter 33. Conducting a Direct Action Campaign | Section 20. Advocacy Over and For the Long Term.* [Online]
Available at: https://ctb.ku.edu/en/table-of-contents/advocacy/direct-action/long-term-advocacy/main
[Accessed 16 March 2024].

Elliston, B., Diesendorf, M. & MacGill, I., 2012. Simulations of scenarios with 100% renewable electricity in the Australian National Electricity Market. *Energy Policy,* Volume 45, pp. 606-613.

Fan, S. & Hyndman, R. J., 2012. Forecasting Electricity Demand in Australian. *IEEE Power and Energy Society Meeting.*

Government, N., 2023. *NSW Climate and Energy Action.* [Online]
Available at: https://www.energy.nsw.gov.au/nsw-plans-and-progress/major-state-projects/shift-renewables/renewable-energy-nsw
[Accessed 16 March 2024].

Korstanje, J., 2023. *How to Select a Model For Your Time Series Prediction Task [Guide].* [Online]
Available at: https://neptune.ai/blog/select-model-for-time-series-prediction-task
[Accessed 16 March 2024].

McCarty, J. P., 2001. Ecological Consequences of Recent Climate Change. *Conservation Biology,* 15(2), pp. 320-331.

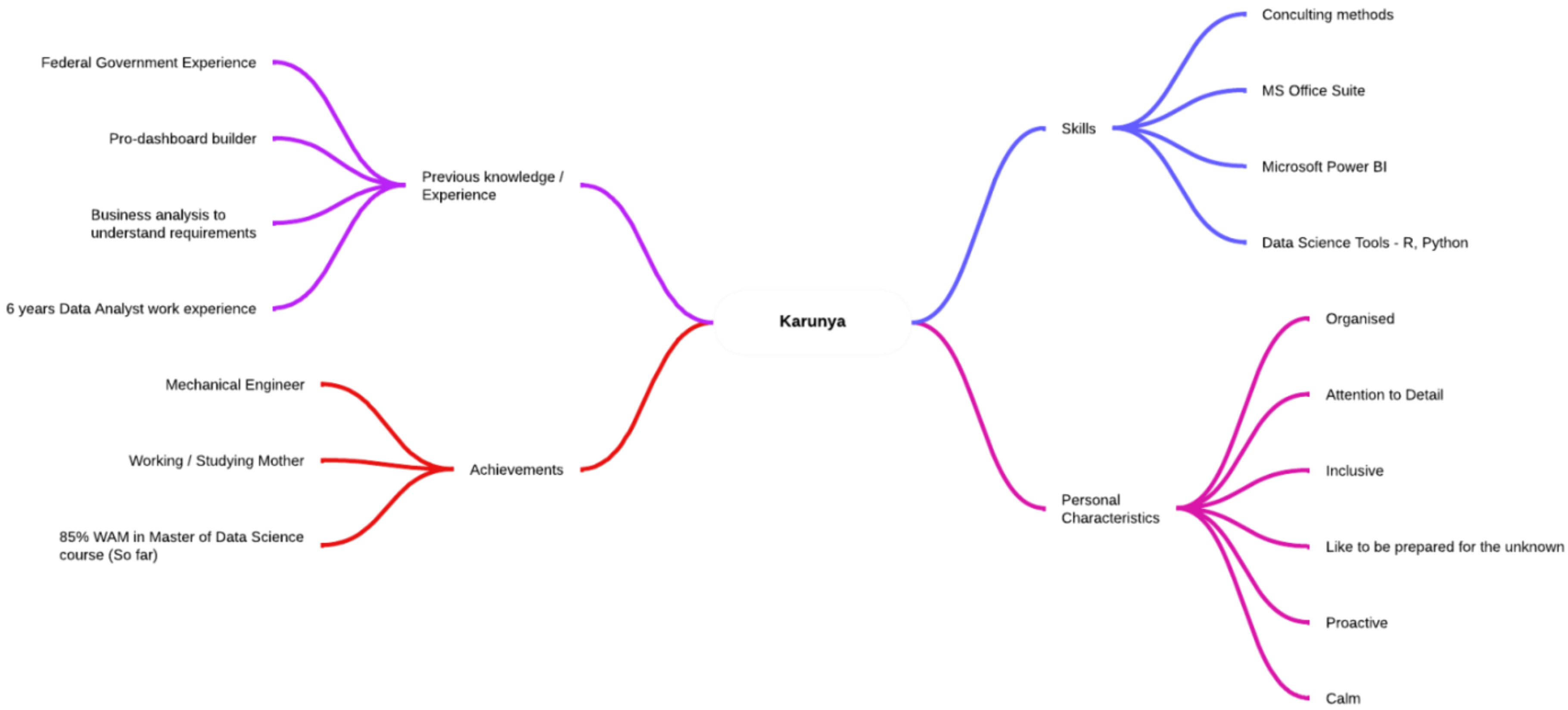
Nguyen, K. & Berry, A., 2019. *Short-term regional electricity demand forecasting,* NSW: CSIRO.

Panesar, S. S. & Wang, W., 2003. Electricity Demand Forecasting Using Neural Networks. *Springer, Berlin, Heidelberg,* Volume 2690, pp. 826-834.

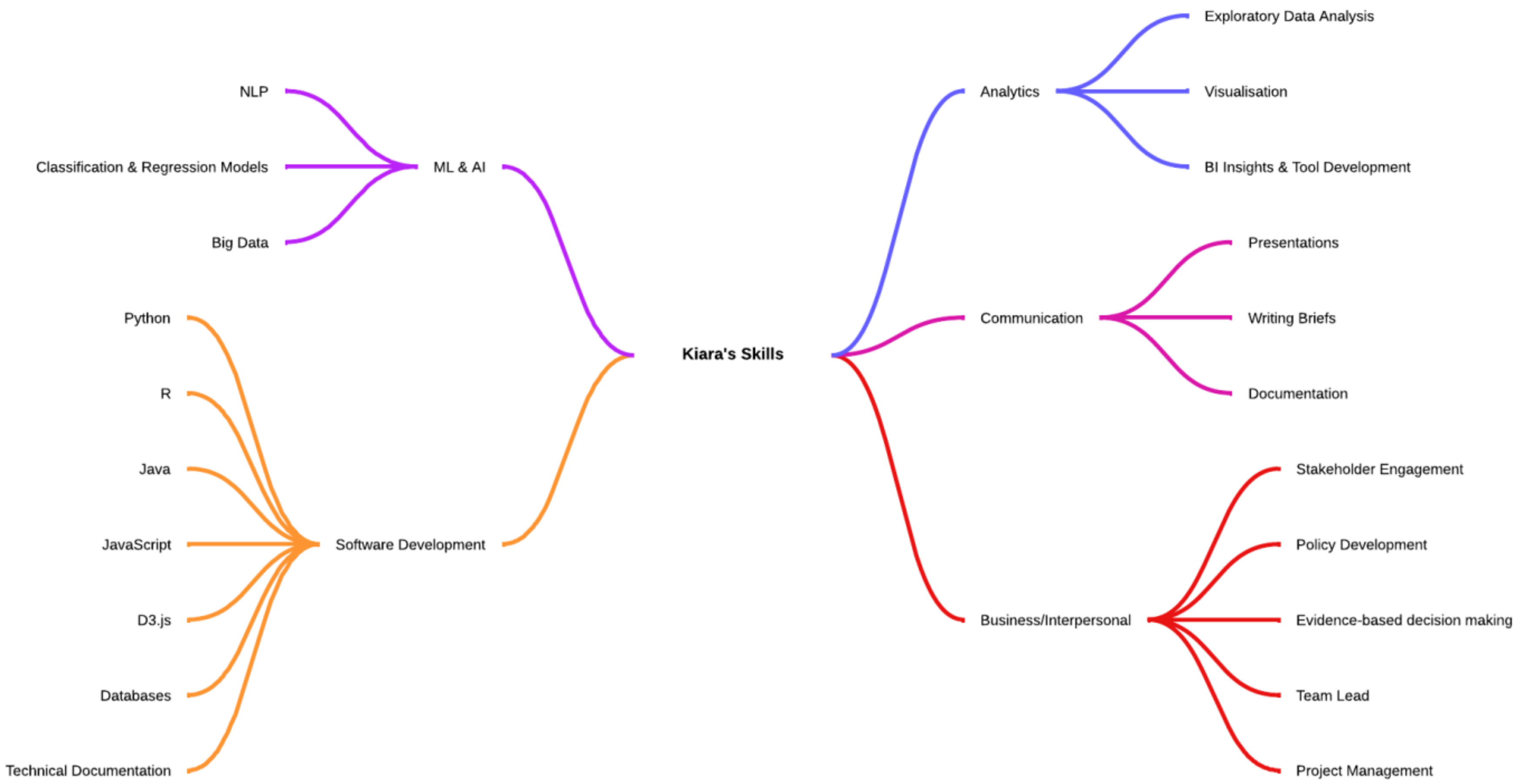
State of NSW; Department of Planning, Industry and Environment (DPIE), 2019. *NSW Electricity Strategy.* [Online]
Available at: https://www.energy.nsw.gov.au/sites/default/files/2022-08/2019_11_NSW_ElectricityStrategyOverview.pdf
[Accessed March 2024].

7 Appendix

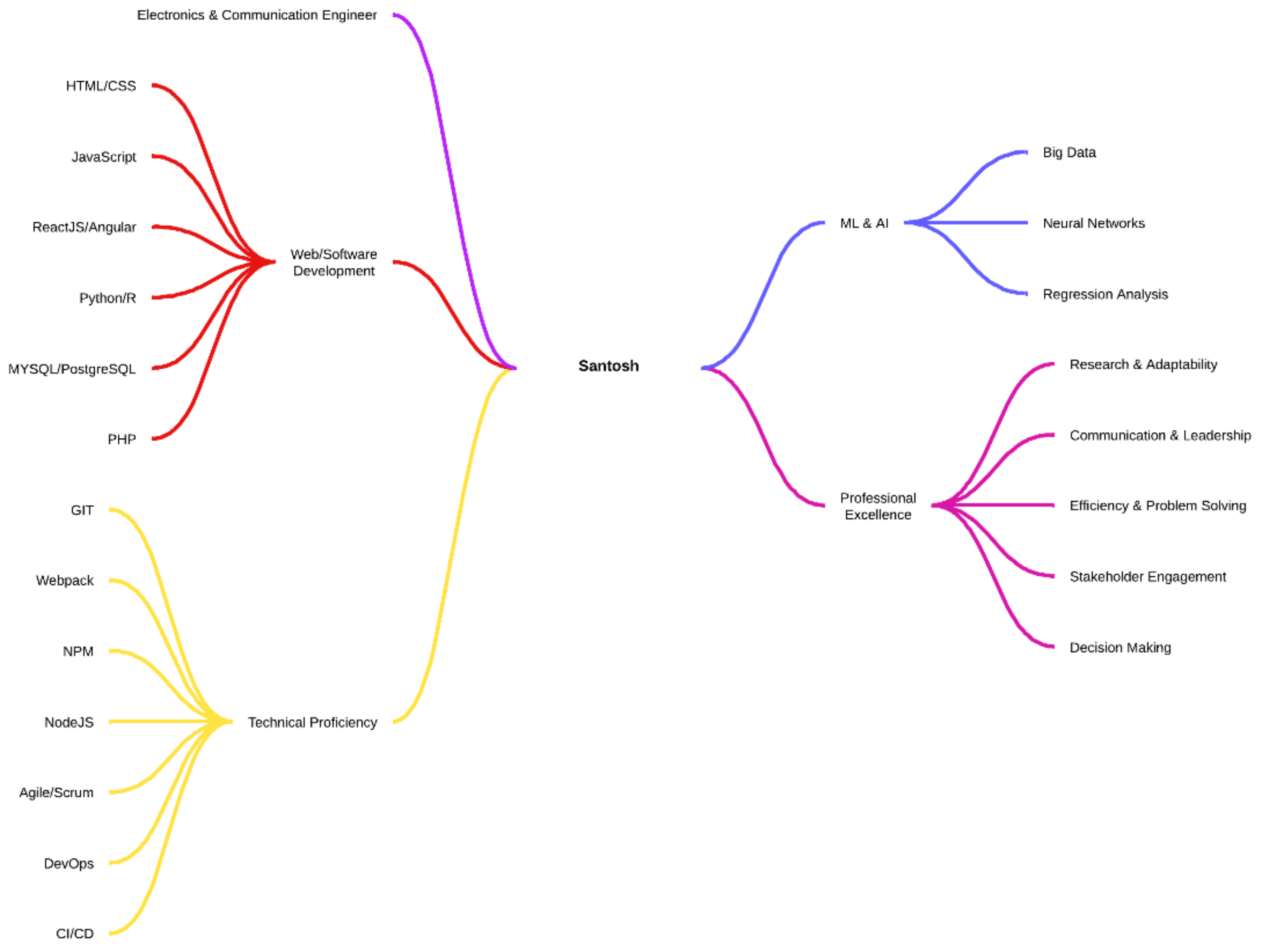
Karunya Skills mapping



Kiara Skills mapping



Santosh Skills mapping



Mathew Fraser Skills mapping