IT INDUSTRY PROJECT - PROJECT REPORT

BUDGET UTILISATION - Data Science Project

Team Techmates

1. **Drashti Nayak (n10599568)**
2. **Roberto Carlos da Silva junior (n10374647)**
3. **Harshal Majithiya (n10550453)**
4. **Sarjak Tapodhan (n10553916)**
5. **Sing Yin Chan (n9317007)**

Project Supervisor: Dr Venkat Venkatachalam

Industry Partner: Leap in!

Industry Supervisor: Jane Sheehy (CTO)

This project plan is submitted to the Science and Engineering faculty in partial fulfilment for the Master of Information Technology degree.

The plan intends to deliver a machine learning predictive model for industry partner Leap In, which will help the organisation continuously track member activities and, based on pattern, provide targetable actionable recommendations.

IFN711 | Industry Project Assessment 2 | Rubric: Final Project Report & IT Artefacts

Criteria	High Distinction	Distinction	Credit	Pass	Fail
Project Outcome and success • *Value* • *Useful* • *Complete* 25 marks	Demonstrates a sophisticated understanding of the project context including the nature of the business or organization and the sector within which it operates The project goals are listed and explained so that their impact is clear. There is a clear and professional assessment of the extent to which the goals were achieved, and explanation provided as to how the assessment was achieved including any metrics use for evaluation.	Demonstrates a high understanding of the project context including the nature of the business or organization and the sector within which it operates The project goals are listed and explained clearly but lack sophisticated engagement and depth of understanding. An evaluation of effectiveness is clearly presented but lacking depth of engagement and inclusion of logical rationale or metric.	Demonstrates a good understanding of the project context including the nature of the business or organization and the sector within which it operates Project goals are listed and an explanation is provided but is missing clear articulation of their impact and value. A general assessment of success is included with limited depth and no inclusion of rationale or evaluative metric.	Demonstrates an understanding of some significant aspects of the project including the nature of the business or industry. The project goals are presented but not explained with depth or detail. Most information is pertinent to the project. A simplistic assessment of success is included with limited, or no rationale or evaluative metric included.	Does not display an understanding of the project correctly, significant aspects missing. Reasoning is deficient. Information is not relevant or flawed. Project goals are not articulated, incorrect or unprofessional in engagement
Project Progress & Reflections • **Critical Analysis** • **Problem Solving** • **Efficiencies** • **Effectiveness** 25 marks	• The project increments were executed completely as outlined in the Assessment 2 Project Plan • All increment level evidence on the project progress is included using accepted tools (Trello images, burndown charts, Gantt Charts and retrospective records) • Revisions to the Assessment 2 Project Plan are discussed and justified professionally. • Relevant experiences related to issues your group faced are critically	• The project increments were executed mostly as outlined in the Assessment 2 Project Plan • Most increment level evidence on the project progress is included using accepted tools (Trello images, burndown charts, Gantt Charts and retrospective records) • Revisions to the Assessment 2 Project Plan are discussed in detail and justified • Relevant experiences related to group level issues are properly analysed, presented with original resolutions	• The project increments were executed partly as outlined in the Assessment 2 Project Plan • Some of the increment level evidence on the project progress is included using accepted tools (Trello images, burndown charts, Gantt Charts and retrospective records) • Revisions to the Assessment 2 Project Plan are described with limited justifications • Relevant experiences related to group level issues are analysed,	• Most of the project increments did not deliver meaningful outcomes • The increment level evidence on the project progress is not relevant or not included from the tools (Trello images, burndown charts, Gantt Charts and retrospective records, “Done” lists) • Limited Revisions to the Assessment 2 Project Plan are mentioned	• No useful outcomes were delivered in all increments • The increment level evidence on the project progress is not included from accepted tools (Trello images, burndown charts, Gantt Charts and retrospective records) • Revisions to the Assessment 2 Project Plan are not covered at all

	analysed, presented with insightful resolutions		presented with basic resolutions	• Relevant experiences related to group level issues are not discussed	• Relevant experiences related to group level issues are not discussed (no group level reflections)
IT artefacts Delivered to the customer • *Quality* • *Professional Standards* **40 marks**	The artefacts and its subcomponents submission fully match with the scope agreed with the client Those are delivered to with high quality standards and well above normal expectations of the industry partner or tutor The IT artefacts are indistinguishable from one produced in a similar time frame by an experienced professional team.	The artefact submission encompasses much of the scope agreed with the client or tutor but falls short in several important areas, with key components either not attempted or completed, or completed well below a professional standard. Overall, the artefact reflects work of a good standard for the work successfully delivered, while remaining below the level as those in the higher-grade band. The work generally reflects a standard at normal expectations of a team of Master of IT final year students.	The artefact submission broadly reflects the scope agreed with the client, but some aspects of the agreed deliverable may not have been completed or may not have been completed to a professional standard. Overall, the artefact reflects work of a near professional standard for the work successfully delivered, without perhaps reaching the same level as those in the higher-grade band.	The artefact submission encompasses some of the scope agreed with the client but falls well short of an acceptable deliverable for the project. Key components are either not attempted or completed or completed poorly. Overall, the artefact reflects work of a relatively weak standard for the work successfully delivered,	The artefact submission encompasses little or none of the scope agreed with the client or tutor None of the agreed components of the artefact have been delivered, and those mentioned are of poor quality. Key components are either not attempted or completed or completed poorly.
Communications • *Clarity* • *Conciseness* • *Correctness* **10 marks**	The report is consistently professional in tone and structure and addresses each of the listed requirements in great detail No errors of grammar or structure. Report is organized to aid understanding, and this is assisted by the layout and formatting.	The report is generally professional in tone and structure and addresses each of the listed requirements in detail Very limited errors in grammar or structure. The report is well organized, and the layout and formatting are well chosen. The standard of writing is above expectations	The report is professional in tone and structure but lacks some detail in a small number of the listed requirements. There may be frequent and occasional errors of grammar or structure. Organization and layout remain good, The standard of writing meets the expectations of this level.	The report does not meet expectations, and the coverage is deficient in several the listed requirements. Grammar and structure are variable but are usually ok. The organization is deficient, but some effort has been made to structure and format the document.	The report doesn't meet the requirements set out in the brief. Sections missing or poorly covered. Meaning unclear as grammar and/or spelling contain frequent errors Disorganised or incoherent writing

		The standard of writing exceeds well above expectations	
		The standard of writing may lie below the expectations at this level	Structure either absent or incoherent and the standard of writing may be well below the expectations at this level

Execution Summary:

Our industry project focuses on Leap In, a non-profit organisation that helps disabled people manage their budget plan. The budget plan is based on NDIS, which is the National Disability Insurance Scheme. Currently, 52,000 dollars were allocated to each member annually. However, the members-only use 68% of the overall budget. Hence, there is a need for data analytics to improve budget utilisation and increase the company's performance. Our proposed solution is a data analytical approach (reports and dashboards) to create an expert user system functional to see insights about budget utilisation from the segmentation process with machine learning techniques. Our project's goal was to determine if members were overspending, underspending, or on track with their budget. However, to analyse the data, our team processes the data with different machine learning techniques. Logistic regression, decision tree and the Neural Network were introduced in the project. Our main approach was the logistic approach, which gives accurate outputs while producing impartial conclusions with lower-level variations. However, our team could not utilise the automatic cleaning technique since there were so many ambiguities in the data, so we had to manually clean it, which took a long time and put us behind the sprint plan. Our model can predict the current state of the members, but it cannot predict the future.

Our approach for the project was DSDM which vastly helped us with reviewing our results with initial requirements as it needs much more experimentation than a direct solution. We also tried and tested multiple different ways to gain insights from datasets and methods of modelling. Even though we were not able to fulfil all requirements of our client, we were able to give a deeper insight into their data and available techniques and methods which can be included in the modelling and business processes of Leap In. It also gave us immense industry experience and how to accommodate clients requirements and manage them around the time variable. Fortunately, we didn't face any shortage of cost and resources. However, we did have to face risks of team member absence and unfinished jobs left out which at last were handled with the support of each other. At the end not only our perception towards visualising data changes and also increased our abilities to work in a team with responsibility, consistency, reliability, honesty, communication and dedication towards the project.

Table of Contents

1. Analysis
2. Design
3. Outcome
- **3.1. Testing and Quality Control**
- **3.2. Quality Assurance**

4. Group Reflection

Appendix A Glossary and Abbreviation

Appendix B Gantt Chart (Initial and Final)

References

1. Analysis:

We were provided with four different datasets to analyze. The dataset was about budget plan, member details, completed claims and active claims which showcases member's spending activity. Firstly, we analyzed the budget plan dataset and member support dataset in which numerical and categorical variables were further preprocessed. After handling missing and duplicate values in both datasets. In the budget plan dataset, we calculated the duration of the total month of the membership plan and in the member support dataset, we calculated the age groups on the basis of date of birth. This gives us the opportunity to analyze different age groups and the duration of their spending activity. We made assumptions in the data preprocessing stage which are defined further. First, we only considered budget plan observation where budget status is confirmed. Second, we differentiate budget plans into completed plans and active plans.

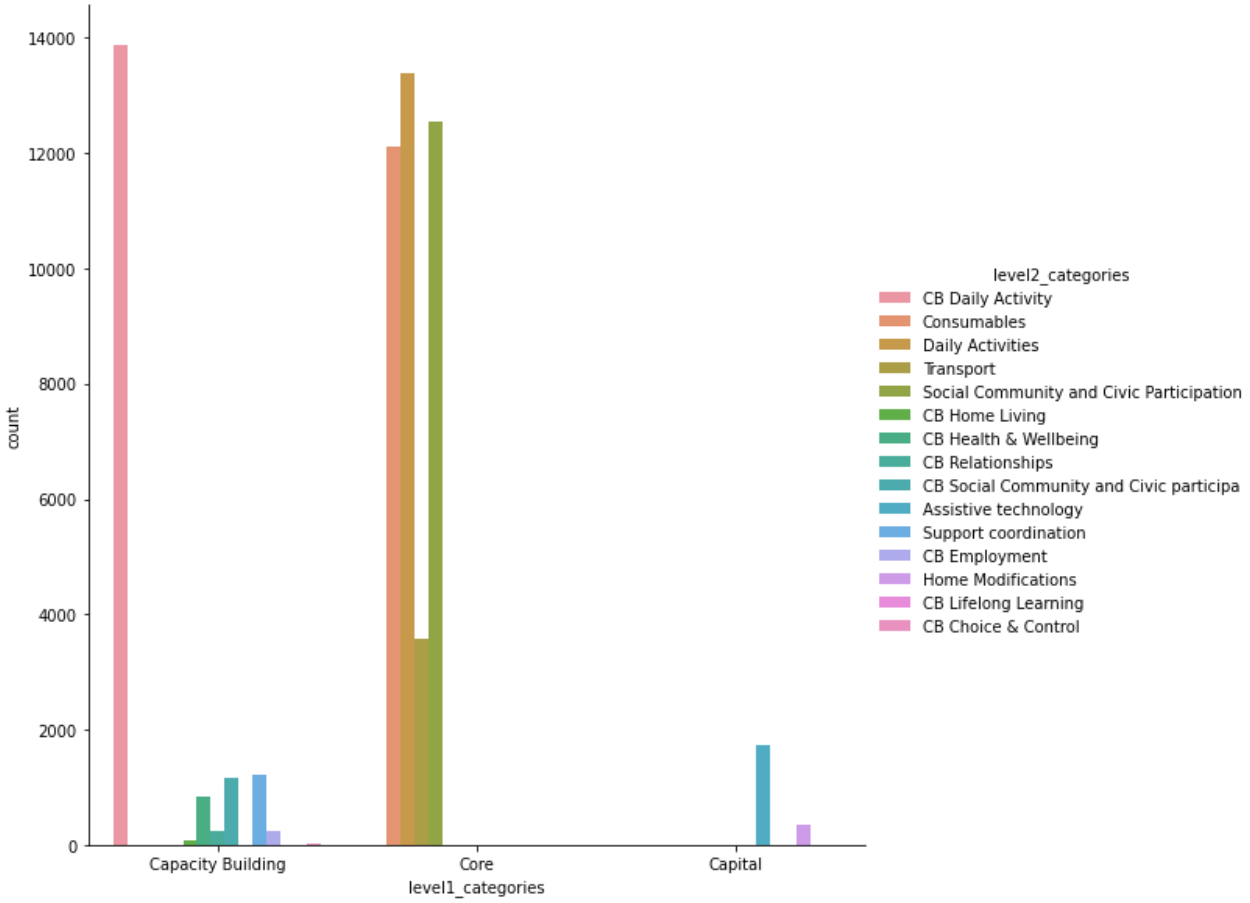


Figure 1 Categorical frequency of spending of members based on two levels of service distribution.

Third, the budget plan also showcases the categories in which the member spends the most. Categories are defined into two levels, the first level contains three categories that are capacity building, core and capital which is further classified into 15 categories in level two. Most of the members spent daily activity in capacity building. The three top used services in the core category are daily activity, social community and civic participation and consumables services. Members

have not spent much in the capital category but mostly spent on support coordination service. Further, we also observed members have spent very least in choice and control, lifelong learning, assistive technology, CB home living, CB health and well-being and CB relationship which can be improved by the Leap-in organization. Later we separated the budget plan data into the completed plan and active plan based on plan_status which we further separately merged with the claims dataset. In the member supported dataset, for the gender category variable, we imputed missing values using the "pad" method. We considered very remote zones into remote zones and we found more than 80% were found in the ACT_NSW_QLD_VIC zone. In gender variables, we discarded other categories and almost 60% of members fall in the male category. We have also separated member data based on active plan and completed plan and further merged it with budget plan data and claim data.

The other two datasets are about the invoices of member's spending activity and the claim process of active plans and completed plans. In both datasets, we have only considered invoices in which invoice_state is either fully paid, short paid, part paid, fully refunded, partially approved, short payment approved, partially refunded or all approved. Then we further selected invoices for whose claims_state was paid and processed_status was successful. It is necessary to consider the claimed amount and not the funded amount because if the claim is not successful then the member cannot access the funded budget and has to pay from his/her account. Hence, funded values are neglected compared to claim values and only successful claims amount is analyzed to understand the spending status of the member. The depth of our analysis can be seen in Tableau Dashboard attached with the artefacts.

Machine Learning stage: The final dataset for machine learning, we considered the budget plans that are **Active** and invoices that are **Paid**. Also, we have considered ***only level 1 services***.

Calculations carried out:

- Total budget duration (in months) further differentiated into Months passed and Months Left.
- Total Budget based on Funded amount per level, Allocated amount per level, Allocated amount should be spent by a month and Allocated amount must have been spent by now.

The Machine Learning model simply follow the following rule:

- On track: If spending_value(x)

$$0.8 < x \leq 1.2$$

- Else Underspending/Overspending

2. Design:

As a first step towards designing the system, the LeapIn scheme process needs to be identified. Image 1 was created to help in a better understanding of this scheme. An NDIS participant, when becoming a Leapin member, will have his data analysed to look for patterns that link his necessity to other attributes such as location, life stage, nature of the disability, available funds, and category of spend. This analysis will produce a plan that will guide users to improve their budget utilization by identifying thin markets, where providers may extend their services to support members with budgets that they can't spend. Leapin creates these plans estimating budgets to be spent into categories. That will be used to pay users claims for services by invoices, which is shown in Image 2.

Image 1 NDIS member budget utilisation process.

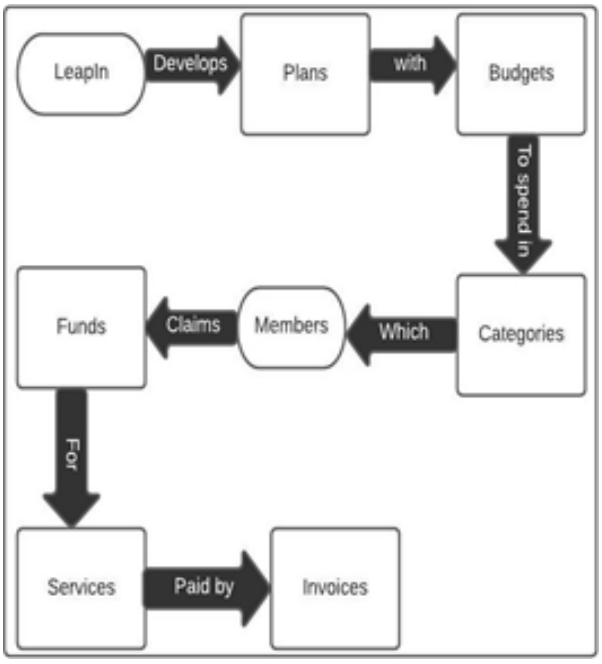


Image 2 Leap In organisation process for handling member's budget plan

LeapIn has all its member's data stores in the MySql database to manage all the processes from when an NDIS becomes a member to processed invoices. We followed a systematic approach based on Data Analysis and Analytics which further differentiated into data cleansing and pre-processing, data analysis and manipulation and Tableau result insights in one way and start working on machine learning framework another way where we follow feature engineering, Dataset Splitting, Evaluation, Hyper-parameter Tuning and Prediction measures. In the first stage, the scheme explained below was understood as a database scheme as shown in Image3, in which the data stored in MySQL tables would be converted into three.CVS datasets, that are:

1. Members.csv: Regarding members' personal information, such as date of birth, first and second name, gender, and disabilities which will help give Leap In an internal view regarding its clients' characteristics and relevant spending interest.
2. Plans_Budgets.csv: Regarding plans and budgets of these members, this layout overall budget plan description with budget amounts and plan duration.
3. Invoices_Claims: Regarding their claims and invoices which are separated into two datasets active claims and completed claims.

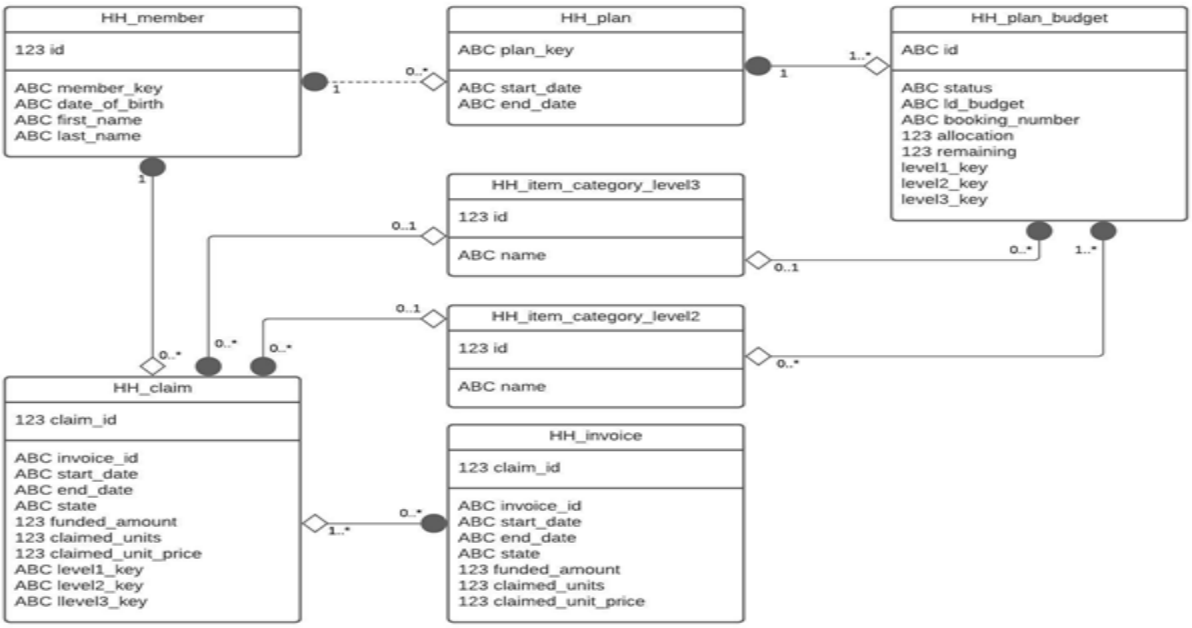


Image 3 Dataset management process Design

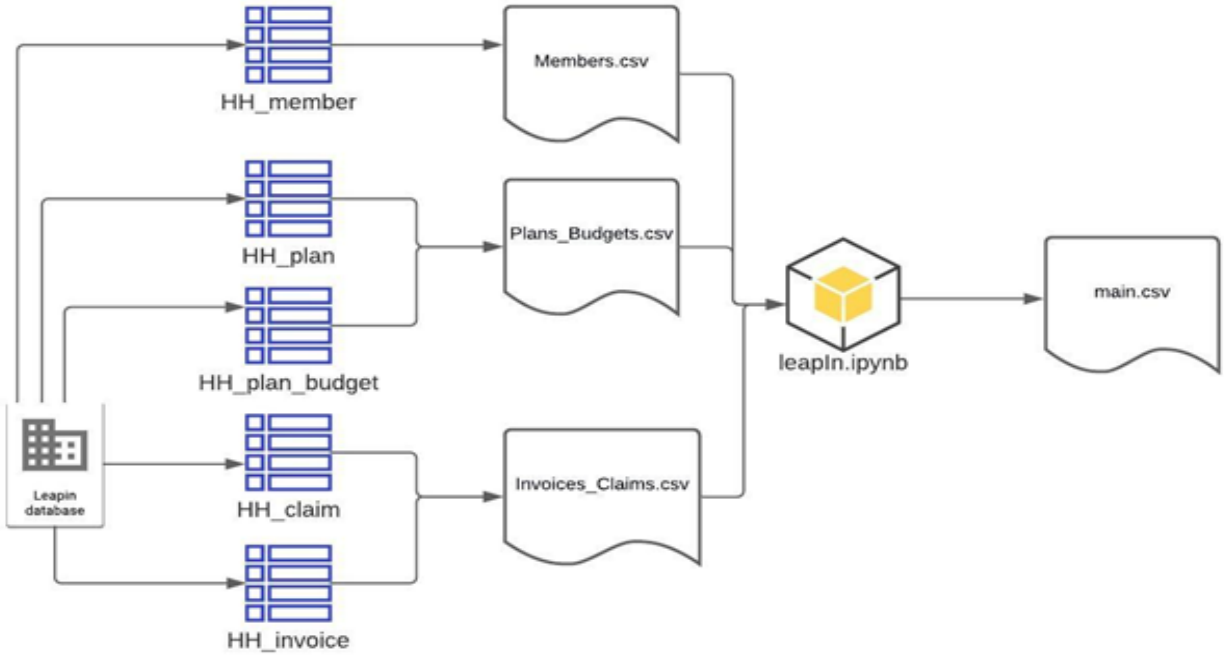


Image 4 File management design of our project

In Image 3, the design of how we handled each dataset and combined it into one meaningful data frame requires us to predict a member's spending_status. Initially, these datasets are passed by cleansing throughout a Python routine called leapin.ipynb as shown in Image 4. This cleaning will

drop or replace the missing values, replace each NaN with a scalar value, or fill forward or backward to detect and correct corrupt or inaccurate records from the datasets, so they were merged turned into a big dataset called main.csv to be used in the two next stages.

In the second stage, the data from main.csv is used to segment members who have underspend or overspent from clustering by Python library Sklearn. To do so, the system routine considers the last plan active of every member, and calculated much each member had to spend along with his last plan in all levels, and how much of it was really spent, which by clustering was separated members who spent more than their allocated value labelled Overspent, who spent less than their allocated value labelled Underspent, and those who spent exactly how much he should label On_Track.

Name	Allocated_amount	amount_spent
Member 1	1000	1100
Member 2	1000	500
Member 3	1000	1000

Image 5 Example of member's spending activity.

For instance, Image 5 shows member 1 who has spent $100 more of he should have spent, this user will be labelled "Overspent", Member two however has spent 500 less than this allocated amount, so he will be labelled "Underspent", whereas Member 3 spent the same amount allocated to him, what makes him "On_Track".

How we found the solution was keeping an understanding of processes such as in active plans how much was the allocated amount, how much was the amount spent from the first day of the active date until today, and the plan's length in months. The allocated amount was split in plan's length which would give the amount a member should spend per month. For instance, Member 1 is engaged in a 12 months' plans, which is allocated $12000, doing allocated amount per plan's length means the member has 1000 to be spent per month, and if his plans have started in January and today is related to Abril means he is in his 3rd month of the plan and he should be spending $3000 (allocated amount per month/month related) until today. Continuing the example instead has spent $3000 the user spent $3300, which means he overspent his amount by 20%. Logistic regression is so used to be segmented by "Underspending" whether a member spent below 80% of its budget, "Overspent" whether a member has spent above 20% of its budget and "On_Track" if this value is zero.

3. Outcome:

We were able to fulfil all Must's, however, failed to fulfil Should and Could which is a recommendation and review system due to lack of time and data being completely raw, unrelated, dummy and fuzzy which did not support the features for creating a recommendation system. We also accomplished function requirements that are:

1. We are able to identify members spending_status by analysing spending activity and applying mathematical calculations as stated in the design phase. The status stated as under or overspent vs spending on track based on past data.
2. We are also able to identify members differentiated under different labels like age, zone, gender and their spending nature and inability to spend in different categories.
3. We are also able to apply this exploration into a working model which can state current member's spending position during their budget plan timeline.
4. This requirement failed due to time constraint, issues in team management and handling minor anomalies in data which made the workings of the dataset unbalanced. Hence, we couldn't fulfil the future prediction capability of member's spending status in the current project timeline.

3.1. Testing & Quality Control:

Our project was focused on providing the outcome of the members if they were Overspending, underspending or were On-track of spending their budget. From the insights we received from the Project owners, most of the members were Under-spending their budget. Our main approach was Logistic regression because of multiclass variances and our output is a categorical variable. Additionally, Logistic Regression uses different methods for estimating the parameters, which give comparatively better results that are unbiased with lower-level variances while providing discrete outputs. We have also compared our approach with the Decision Tree models and Neural Networks to observe which one has a better predictive capability. The results proved Logistic Regression was a better Machine learning technique with 99% accuracy.

For testing we followed a machine learning framework:

1. We used a method in which we split the data into a Train and Test set.

2. Then we evaluate the model.

3. Then we work on the parameter tuning.

4. After that the model predicts.

In Logistic regression, we split the data into an 80-20% ratio. As the result is a categorical variable that showcases the spending status of members, we used an Ordinal Encoder to transform the variable. When we evaluated the model we found the model was overfitting hence we used the SMOTE 'Synthetic Minority Oversampling Technique' for Class-balanced Tasks, which is a type of data augmentation.

In the Decision Tree, after applying the Ordinal Encoder, we observed the model prediction on both Gini and Entropy criteria. Entropy measures a slightly higher prediction than Gini for two depths of the tree. We kept the Entropy criterion because it generates more Balanced Trees. In the evaluation stage, we found that max_depth =4 generates a more stable result. We also considered experimenting with different random state values, but we found no change in the prediction. We also tried applying the Minimal Cost–complexity Pruning algorithm in the parameter tuning stage. It increases the training error but reduces the testing error which makes the Decision tree more adaptable. However, we found the prediction accuracy to be 86.45%, the same for both the algorithms so pruning algorithm is not necessary.

In Neural Network, we have used MLP classifier and Stratified K-Fold method to split the data. Stratified K-fold cross-validation will enforce 5 splits of data to balance class distribution. As a result, the overall accuracy of the model is 77.65%. We have not been able to further tune the model due to time constraints. The probabilistic Neural network approach can be investigated in future. And we were also not able to explore complex activation functions.

Logistic Regression	99.96%
Decision Trees	86.45%
Neural Networks	78.00%

Image 6 Highest Accuracy achieved by each model.

Additionally, we also tried to create a rule-based prediction model which doesn't require a machine learning process. But after the data pre-processing stage, we found spending activity per person is not enough and requires observation for at least one year. In Logistic Regression, when we considered the whole data set after encountering all the uncertainties in the variables, we achieved a 77% precision only. But when we used only the spending limit for predictions we achieved a 99% accuracy, which is quite substantial.

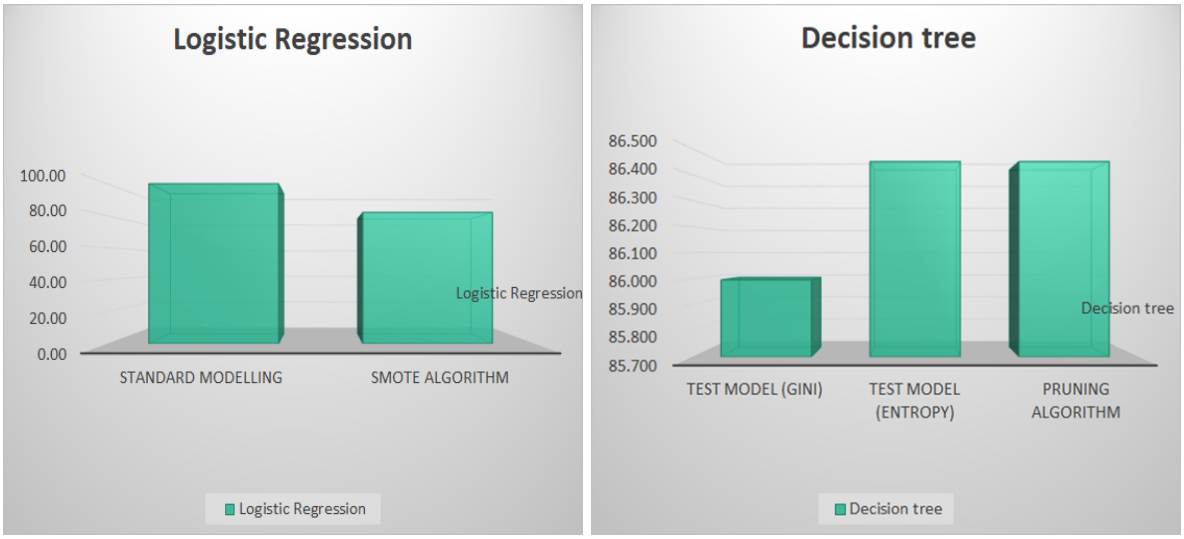


Image 7 & 8 Logistic Regression and Decision Tree predictive capability

3.2. Quality Control:

For quality control, we mainly focused on the Non-functional key requirements for this project. There are three main aspects which we focused on.

1. Cloud service: We used Google collab as our primary coding environment as it is totally cloud-based. Rather than opting for Jupyter notebook as our development for our coding platform because everything is saved in the local machine of the system. Google collab made it really easy for us to access the file from any device with a simple Google log-in. All of our Google collab notebooks are saved on our Google Drive account.
2. Performance: This is kind of a no-brainer to choose Google Colab over the local Jupyter notebook because google research lets us use their dedicated GPUs and TPUs for personal machine learning assignments. GPU and TPU acceleration does make a really big impact even for small projects. Especially speaking from personal experience the GPU from Google resources works way better when it comes to neural network operations compared to the local Jupyter notebook.
3. Security: One of the advantages of using Google Colab is its privacy features. Users can only access the notebooks through the Google Log-in, where it sends you a unique key which you paste in the Google collab and only then you have access to the data, this shows how much secure Google collab is compared to the local Jupyter notebook.

4. Group Reflection:

4.1. How we applied knowledge we gained from our major units:

Our knowledge from IFN509 Data Mining and Exploration vastly helped to discover a solution for this project. The unit explained to us to learn the basic concept of statistics for data exploration and gain insights and understanding the machine learning framework to gain results for different data types using different modelling approaches. We were also greatly benefitted from unit IFN619 Data Analytics for Strategic Decision Making which enhanced our perception towards data handling and story-telling.

4.2. How we collaborated to address the problems and issues encountered during the execution:

We can state that problems that we faced during the execution were not technical but more internally of how our team used to work. As many members having jobs, we needed to make changes in weekly meeting every week to accommodate most members. However, many times absence made a great deal for the tasks that need to be completed and had to send back to the backlog. We had a time when two of our team members were going through personal difficulties and those times were not only hard for them but the whole team but we learned the work of teamwork and support during that duration.

4.3. Reflection about learnings on new tools and technologies:

We used google collab and tableau for the first time. Tableau is a great tool to showcase business insights with each operation by just dragging and dropping but it requires hands-on experience to make the most out of it. And google collab literally made our python programming easy to work on and retrieve fast outputs.

4.4. Group-level challenges faced and how we addressed them and necessary actions are taken:

Group-level challenges were higher than project management level challenges. We found ethical issues among our team. We trusted and respected each other's participation in work, but many times we found a lack of honesty, responsibility and fairness among members. However, in absence of a member others extensively supported and managed project tasks. The communication dishonesty within the team greatly affected the quality of work. The majority of the team showed a lack of responsibility and fairness while doing the task allocated or participating fully in the project. We did take necessary actions by evenly distributing the reso of work among the responsible members. We increased our communication methods and talked about the issues with non-responding members. This later in the project, helped us to finish our jobs in time. Our team members from different majors who were unaware of the data science field found difficulty in adjusting, adapting and working on data science concepts. Hence, we separately shared youtube sessions and separate meeting kept to teach them the necessary terminologies.

4.5. Experience in working with an Industry Partner and key learnings of the team:

Experience with Industry Partner vastly changed our view of how corporate culture works. The communication methods, handling client's requirements, changes expected at last minute and how to meet their expectations. It's difficult to stay on track with different working timelines. We learned that what any industry requires is a social skill and staying updated with your work and industry work. This experience also gave self-understanding awareness personally to each of us. Additionally, we learned how to consistently put up with the client's expectations. It resulted in a great learning experience.

4.6. Processes and methods were adopted by our team to handle the changes encountered in the project:

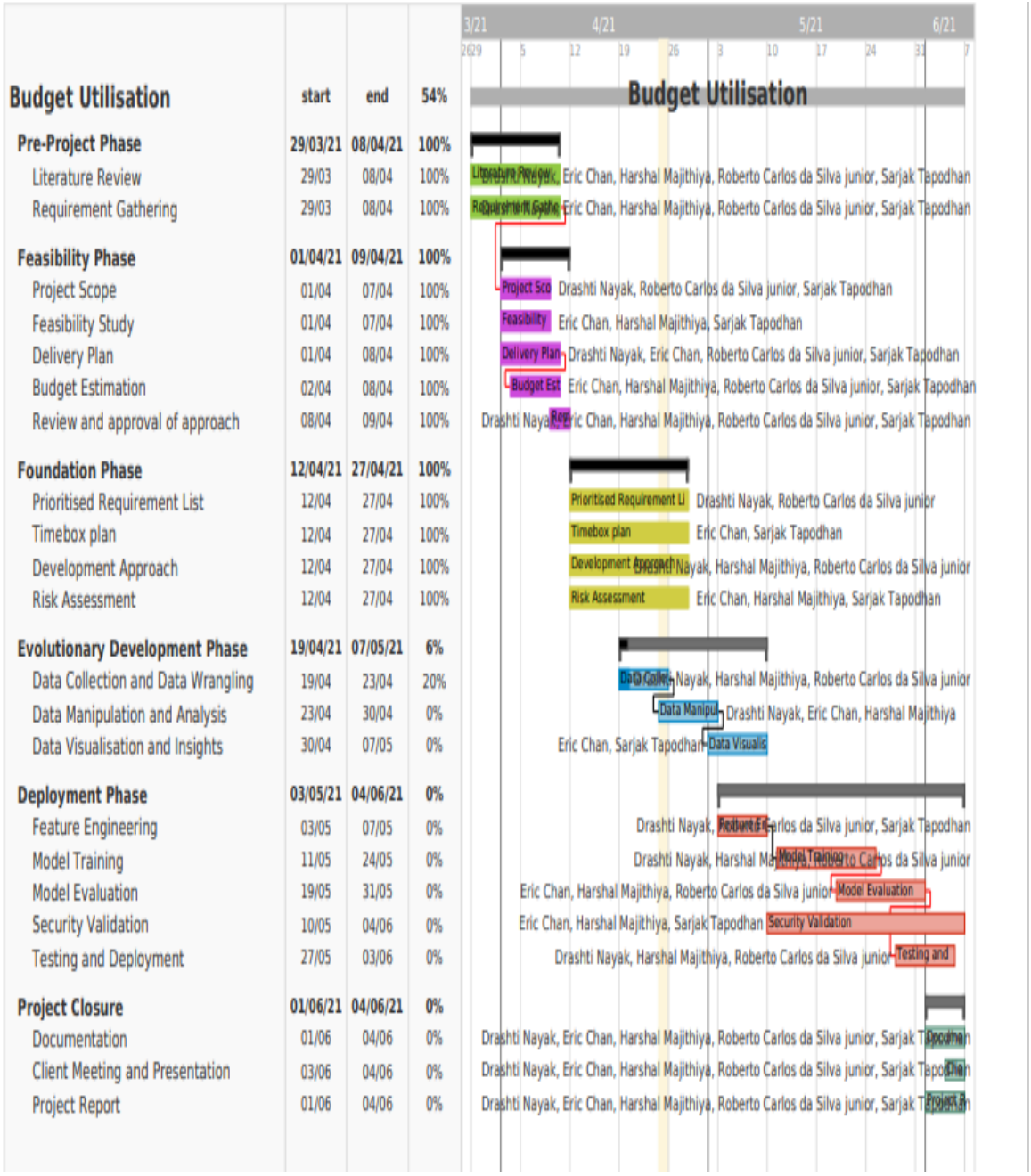
Firstly as data was vastly distributed we chose univariate, bivariate and multivariate analysis in Explorative data analysis which gave us deeper insights into the data. In machine learning, we tried prediction capability with a whole dataset or only required extracted features which showed us another technique that is feature engineering. Working with time series analysis is complex which additionally includes automotive future prediction. Hence, we would further like to learn about the automotive process and the application of other more extensive types of neural networks in similar projects.

Appendix A: Glossary and Abbreviations

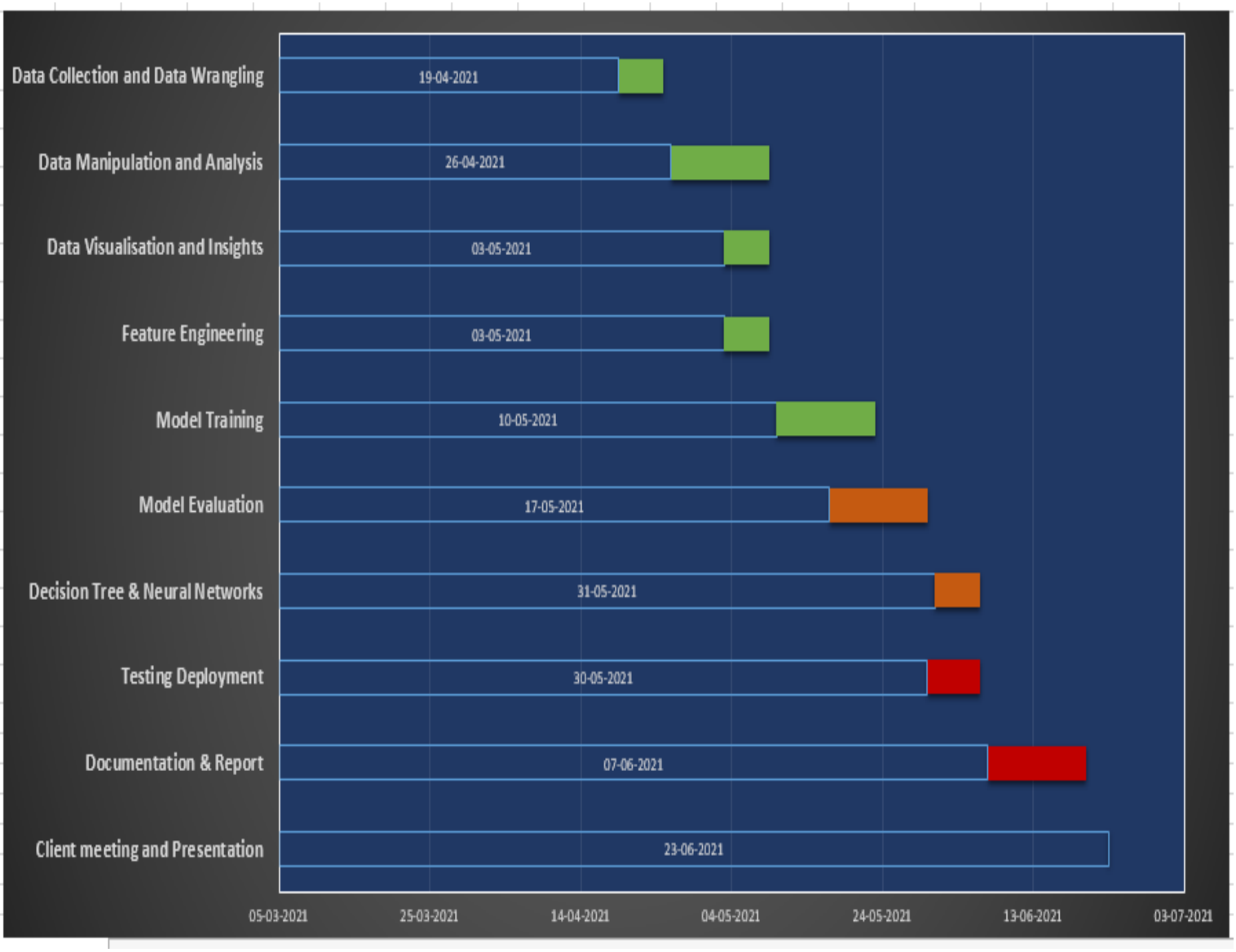
1. Machine Learning (ML): Machine learning is a subcategory of AI which is a model that learns through experience with large sets of data using algorithms.
2. Deep Learning: Deep learning is a subcategory of ML which is inspired by how our brains work. In short, it connects the dots to get relevant information.
3. Data Analysis: Data analysis is a study of past or historical data to get insights and relevant information regarding past environments and how data has emerged.
4. Data Analytics: Data analytics is a study of present data to get predictions of the future based on fact-based data.
5. Data Mining: Data mining is a process of extracting patterns from data by manipulating and applying statistics to it.
6. Data Visualisation: Data visualisation is a process of representing the patterns and information from data mining. So that it can be understood by business teams.
7. Logistic Regression: Logistic regression for multi-class can determine for probability for a scenario based on multi-variables.
8. Exploratory analysis: this analysis used to summarise the results of statistical techniques applied to datasets.
9. Prescriptive analysis: Prescriptive analysis is finding the best solution for a given problem with available data. It is a combination of descriptive and predictive analytics.
10. Accuracy: Accuracy defines how precise a machine learning model gives correct outcomes.
11. GPU: It is a specific processor used to handle large-size data sets and for machine learning.
12. Tableau: Tableau is a data visualisation tool that helps to create visual dashboards.
13. Decision tree approach: A decision tree approach that has a diagram or chart that helps determine a course of action or show a statistical probability.
14. Neural network approach: Neural network approaches are essentially an extension of the empirical methods with parameter fitting, albeit a sophisticated one. They involve a mathematically based assessment of complex inter-relationships within systems.
15. Hyperparameter tuning: In machine learning, hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm.
16. Smote technique: SMOTE (synthetic minority oversampling technique) is one of the most commonly used oversampling methods to solve the imbalance problem
17. Minimal cost-complexity pruning algorithm: Minimal cost complexity pruning recursively finds the node with the “weakest link”.
18. MLP classifier: MLPClassifier stands for Multi-layer Perceptron classifier which in the name itself connects to a Neural Network.
19. Stratified k-fold method: In stratified k-fold cross-validation, the folds are selected so that the mean response value is approximately equal in all the folds.
20. Google collab: Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.
21. GPU: Graphics Processing Unit) A programmable processor specialized for rendering all images on the computer's screen.
22. TPU: Tensor Processing Unit (TPU) is an AI accelerator application-specific integrated circuit (ASIC) developed by Google specifically for neural network machine learning

Appendix B: Gantt Chart

Initial Plan:



Final execution Timeline:



References

- Anolytics. (2020). *How To Improve Machine Learning Model Performance: Five Ways.* https://www.anolytics.ai/blog/how-to-improve-machine-learning-model-performance/
- Baader, F., & Sattler, U, (2001). An overview of tableau algorithms for description logics *Studia Logica, 69*(1), 5-40
- Bisen, S, V. (2019). *How To Ensure Data Quality For Machine Learning And AI Projects.*Medium. https://medium.com/vsinghbisen/how-to-ensure-data-quality-for-machine-learning-and-ai-projects-c8af1fe18c57
- Burrell, J. (2016). *How the machine 'thinks': Understanding opacity in machine learning algorithms*
- Big Data & Society. https://doi.org/10.1177/2053951715622512
- Centric Consulting. (2021, March 26). Machine Learning: A Quick Introduction and Five Core Steps. Retrieved from https://centricconsulting.com/blog/machine-learning-a-quick-introduction-and-five-core-steps//
- Chabot, C., Stolte, C., & Hanrahan, P., (2003). *Tableau software* Tableau Software, 6
- Dhavel.M. (2018). *How to perform Quality Assurance for Machine Learning models.* Medium.https://medium.datadriveninvestor.com/how-to-perform-quality-assurance-for-ml-models-cef77bbbcfb
- Frameworks for Approaching the Machine Learning Process. (n.d.). Retrieved April 25, 2021, from https://www.kdnuggets.com/2018/05/general-approaches-machine-learning-process.html
- Hall, H. (2020, February 12). How to Identify Scope Risks. Retrieved from https://projectriskcoach.com/how-to-identify-scope-risks/#:%7E:text=Scope%20risks%20are%20uncertain%20events,related%20to%20the%20project%20scope.
- IBM Cloud Education. (2021). *Exploratory Data Analysis.* Retrieved from https://www.ibm.com/cloud/learn/exploratory-data-analysis
- Leap in. (2021). *Top 5 reasons why people don't spend their NDIS budgets.* https://www.leapin.com.au/5-reasons-ndis-budgets-unspent/
- Kansal, N. J., & Chana, I., (2012). *Cloud load balancing techniques: A step towards green computing*
- IJCSI International Journal of Computer Science Issues, 9(1), 238-246.
- Kurek, E., Johnson, J., & Mulder, H. (2017). Measuring the value of enterprise architecture on IT projects with CHAOS research Syst. *Cybern Inform, 15*(7),13–18
- Messenger, S. (2014). *DSDM Agile Project Framework Handbook.* The Agile Business Consortium. https://www.agilebusiness.org/page/ProjectFramework_00_welcome
- National Disability Insurance Agency, (2021). *What is the NDIS?* https://www.ndis.gov.au/understanding/what-ndis
- NDSP Plan Managers, (2020). *Are You Using Your Plan Funds?* https://ndsp.com.au/are-you-using-your-plan-funds
- Public Services and Procurement Canada. (2019). *Scope Management Techniques.*https://www.tpsgc-pwgsc.gc.ca/biens-property/sngp-npms/bi-rp/conn-know/portee-scope/techniques-techniques-eng.html

- **Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches.** ***IEEE Data Eng. Bull., 23*****(4), 3-13.**
- **S. Song et al., (2013). Prescriptive Analytics System for Improving Research Power.** ***2013 IEEE 16th International Conference on Computational Science and Engineering, Sydney, NSW, Australia,*** **(2013, pp. 1144-1145), doi:10.1109/CSE.2013.169**.
- **Science, D. (2020, July 20). 7 Stages of Machine Learning — A Framework - Data-Driven Science. Retrieved from https://medium.com/@datadrivenscience/7-stages-of-machine-learning-a-framework-33d39065e2c9**
- **Snowden, J, D., & Boone , E, M. (2007).** ***A Leader's Framework for Decision Making*****. Harvard Business Review https://hbr.org/2007/11/a-leaders-framework-for-decision-making**
- **Tan.J. (2020).** ***How to improve data quality for machine learning?*** **Towards Data Science https://towardsdatascience.com/how-to-improve-data-preparation-for-machine-learning-dde107b60091**
- **Templatelab. (2021).** ***40 Detailed Contingency Plan Examples (& Free Templates)*****. https://templatelab.com/contingency-plans/**
- **TensorFlow. (2021).** ***Welcome To Collaboratory*****https://colab.research.google.com/notebooks/intro.ipynb**