***COMP3850 Project Deliverable Certificate***

| Name of Deliverable | *Project Plan/Quality Manual* |
| --- | --- |
| Date Submitted | *01 / 09 / 2021* |
| Project Group Number | *6* |
| Rubric stream being followed for this deliverable (highlight one)  *Note: the feasibility study has the same rubric for all streams.* | *SOFTWARE Rubric*  *GAMES Rubric*  *CYBERSECURITY Rubric*  *DATA SCIENCE Rubric* |

We, the undersigned members of the above Project Group, collectively and individually certify that the above Project Deliverable, as submitted, **is entirely our own work**, other than where explicitly indicated in the deliverable documentation.

| INITIALS | SURNAME | GIVEN NAME | STUDENT NUMBER | SIGNATURE *(IN-PERSON OR DIGITAL)* |
| --- | --- | --- | --- | --- |
| ***KL*** | ***Lazenby*** | ***Kurtis*** | ***45610266*** | ***kl*** |
| ***KK*** | ***Kannan*** | ***Keerthana*** | ***45870969*** | ***kk*** |
| ***NS*** | ***Soares*** | ***Nathan*** | ***45382417*** | ***ns*** |
| ***TB*** | ***Bienasz*** | ***Thomas*** | ***45982201*** | ***tb*** |
| ***ES*** | ***Salafranca*** | ***Ella*** | ***45524742*** | ***es*** |
| ***XC*** | ***Chen*** | ***Xinrui*** | ***44871481*** | ***xc*** |

*© Macquarie University, 2021*

**List of tasks completed for the deliverable and activities since last deliverable certificate with totals for each individual team member and whole team**

*(copy individual total row for each member and copy pages if more pages needed)*

| Performed by *(Student Names)* | Duration  *(hrs)* | Complexity  *(L, M, H)* | Name of task | Checked by *(Initials)* |
| --- | --- | --- | --- | --- |
| **Xinrui Chen** | 6 | m | Statement of Purpose/Scope/Definition | es |
| 6 | m | Reviews and Audits, Testing, Tools | es |
| 6 | m | Tracking/Change Management and Tools | es |
| 6 | m | Conflict Resolution/Negotiation | es |
| 6 | m | Standards/Templates/Appendices/Forms | es |
| *XC Total* | **30** |  |  |  |
| **Keerthana Kannan** | 6 | m | Risk Management | xc |
| 6 | m | Resource Management | xc |
| 6 | m | Team Organisation and Structure | xc |
| 6 | m | Timeline | xc |
| 6 | m | Resources Allocated | xc |
| *KK Total* | **30** |  |  |  |
| **Ella Salafranca** | 5 | m | Task/Activities/Phases | kk |
| 5 | m | Process Model discussed/justified | kk |
| 5 | m | Documentation identified/discussed | kk |
| 5 | m | Assumptions | kk |
| 5 | m | Quality Control and Management | kk |
| 5 | m | Communication | kk |
| *ES Total* | **30** |  |  |  |
| Team Total |  |  |  |  |

**DeepAudit Pty Ltd**

**Project Plan/Quality Manual**



**Thomas Bienasz**

**Xinrui Chen**

**Keerthana Kannan**

**Kurtis Lazenby**

**Ella Salafranca**

**Nathan Soares**

02-09-2021

COMP3850 - Computing Industry Project

**Contents**

[**Project Plan**](#_iy3cg7tlqvex) **5**

[Executive Summary & Statement of Scope](#_mypetigdxy3c)A picture containing drawing, clock

Description automatically generated 5

[Risk Management](#_iv63akrn7pyr) 5

[Resource Management](#_2qcjais9ng13) 9

[Team Organisation and Structure](#_3wo4zhpl5dhy) 10

[Project Schedule](#_glgv9bv8oo5y) 11

[Timeline](#_6qfsopu1deqw) 13

[Resource Sheet](#_ytvcukbtybmv) 14

[Network Diagram](#_nj9j1enhs0ei) 14

[Process model discussed/justified](#_8286f7n3d3c2) 16

[Documentation Identified/Discussed](#_9wmmuafwk80x) 17

[Assumptions](#_mjekj2ykiv4t) 18

[**Quality Manual**](#_xehsbcba39yb) **19**

[Quality Control and Management](#_19e90v1uge2e) 19

[Reviews and Audits, Testing, Tools](#_tdj7pf3jvdpa) 21

[Tracking/Change Management and Tools](#_y34bcvz3qkk9) 22

[Communication](#_pq472prwmrk4) 24

[Conflict Resolution/Negotiation](#_g8wzy4y6duk0) 27

[Standards/Templates/Appendices/Forms](#_i769limj6j3) 27

[**References**](#_16kbj8qrko3x) **28**

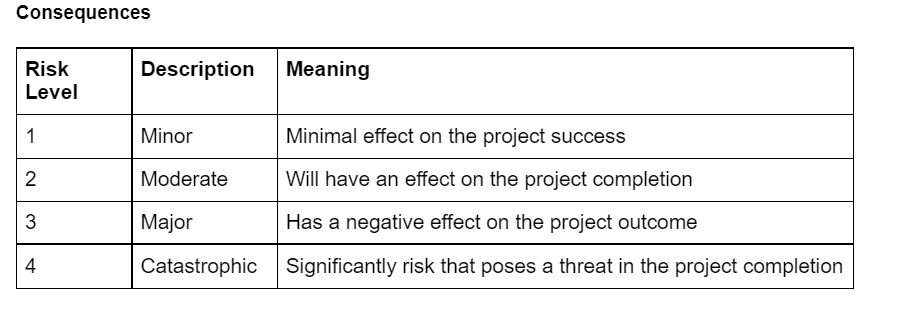
# Project Plan

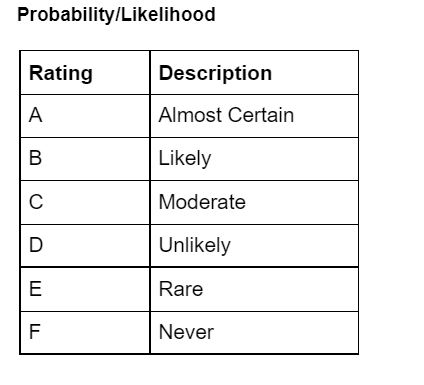
## Executive Summary & Statement of Scope

This document aims to highlight the scope of this project, which is based around machine learning in the financial auditing process. The project plan will analyse the process of implementation of new AI technology to enhance ASIC’s ability to detect poor market disclosure and any risk that may occur.

The purpose of the project is to provide a program which takes in a large amount of financial data and a modelling system that automates the financial auditing process of this data, allowing us to detect fraudulent activity. This will be achieved by using mass amounts of ASX financial data, collected from MorningStar in Excel format. This is then converted to an xls format in order to function properly on JupyterNotebook. The data is then cleaned and input into our AI based modelling system, which trains our models using clustering techniques. Thus, allowing us to detect fraudulent data as outliers will begin to arise as more data is put through the program.

## Risk Management

A crucial part of the project planning phase is to identify all risks, negative and positive, analysing their effect on the project and having a response plan to meet project milestones. Being able to successfully manage risks in IT projects is the central problem for businesses and results in a low success rate for IT projects ***(S.Liu,2014)***. A risk takes place when the project timeline, performance or budget is impacted, therefore it is necessary to identify, analyse, plan and monitor these risks. When a risk is identified, there are to be prioritised and ensure mitigation strategies are put in place.

****

**Risk Matrix**

| **RISK** | **Description** | **Consequence** | **L** | **P** | **Risk Rating** |
| --- | --- | --- | --- | --- | --- |
| Lack of research | As this project involves machine learning and a thorough understanding of financial statements it means a lot of time is invested in researching before beginning to plan the project. | Not obtaining enough information will mean that the project completion plan will not meet the project goal. The end goal cannot be achieved if there is not enough background research done before starting and the team is directed to the wrong path | 4 | B | High |
| Low Quality Data | Good quality data will have the correct information needed to be used in the project. There is a lot of data available for this project but it is necessary that the right data is selected from the correct database and is appropriate to meet project requirements. | Incorrect, irrelevant and less number of data will affect the data preparation time as well as the end result. | 3 | D | Medium |
| Data Storage | This project involves various types of files to be utilised such as documents, articles, datasets and jupyter notebook files. | Not having an appropriate storage solution means there is a requirement to have multiple duplicate files saved on other hard drives and USBs which could cause misinterpretation of files. | 2 | D | Moderate |
| Scope Creep and Sponsor Discrepancies | There are high chances of the scope being unclear and not defined properly. The main cause of this is due sponsor discrepancies, where the sponsor's expectations are understood differently by the project team. This means the deliverables in the end are going to be completely different and misinterpreted. | The project fails to meet the expectations of the sponsor and results in an unsuccessful project. The project therefore, does not achieve the business requirement and the overall project goal that the sponsor is expecting. | 4 | B | High |
| Project Integration | Working as a team to complete a project that cannot be separately and even if they are separately, a problem arises when they are combined yet missing parts to complete the project. | When working together on a big project, it is necessary to break the project into small sections and integrate them together but this means that there are high chances for sections to be missed out or incompleted which could affect the time and overall performance of the project. | 2 | B | Medium |
| Team Commitment | Project team lack of commitment to completing the project on time and properly. This is usually caused by lack of motivation. | Unable to complete the project satisfactorily and not meet sponsor's high expectations for the project. | 2 | E | Low |
| Time Constraint | Underestimating the time that certain milestones will take in the project schedule plan and the project taking longer than expected in one part delaying the others. | This risk could potentially lead to failure to meet the project requirements as well as not being able to complete a project at all. | 4 | B | High |
| Conflict within Project Team | Unprofessional behaviour and disagreements within the team as well as towards the project. | Time is wasted to resolve conflicts and could affect the overall completion of the project. | 3 | E | Medium |
| Health and Safety | Due to the increase in Covid-19 cases in NSW, there is a risk involved where a project team member could contract the virus. | The consequence of the local outbreak could result in the project taking longer to be completed or certain outcomes not being achieved. | 2 | B | Medium |
| Data Science skill set | Project team lacking data science and project skills and failing to successfully use data science tools. | Unable to complete this data science project and not meeting the required project deliverables. | 2 | C | Moderate |

**Risk Treatment Plan (Mitigation Strategies)**

| **RISK** | **Mitigation** | **Rating Before** | **Rating After** |
| --- | --- | --- | --- |
| Lack of research | Allocating enough time to research academic articles and other company websites to gather as much information as possible to assist with gaining a clear understanding of the knowledge required for the project, especially understanding the financial side of the project. | High | Moderate |
| Low Quality Data | Gathering the required datasets from an official data provider and using the same database site for all data collection so the data is consistent. | Medium | Low |
| Data Storage | Utilising cloud storage instead of local storage for data and documents such as Github and Google drive. | Moderate | Low |
| Scope Creep and Sponsor Discrepancies | Having regular meetings as a group as well as meeting with the sponsor to get weekly feedback on the project progression and logging the meeting notes. | High | Moderate |
| Project Integration | On Github having individual folders for the project to complete their parts and a combined file to produce the complete project and constantly updating readme checklist after completion with comments. | Medium | Low |
| Team Commitment | Evenly splitting the work between the team so there is no disadvantage for certain members as it may cause them to lose motivation to complete the project if there's a large workload. | Low | Low |
| Time Constraint | Having a proper schedule and shorter milestones to make sure all tasks are completed within the given timeline. | High | Medium |
| Conflict within Project Team | A clear understanding of the professional behaviour and positive attitude required when working on a project as a team. | Medium | Low |
| Health and Safety | Following lockdown restrictions put in place and staying safe. | Medium | Moderate |
| Data Science skill set and Tools | Learning new tools and improving skill sets through online tutorials from past units or youtube and other online learning platforms. | Moderate | Low |

Most of the project risks can easily be mitigated through various strategies given in the table above. These strategies can only be implemented if the team collaborates well together and make sure they create a successful project plan to follow as well. As a general rule of thumb, it is important that the project team clearly interpret the expectations of the Sponsor and utilize their knowledge and understanding to come up with a solution as a team.

## Resource Management

**Hardware:** Throughout the project timeline, there are requirements for hardware resources that are essential to start and complete the project. These hardware resources include:

* Personal Laptop or PC
* Computer Data storage or USB
* Keyboard
* Mouse
* Monitor
* Headphones with a microphone

**Software:** The project requires the use of multiple softwares to be used and these are not just for the Data Science project but also to showcase the other deliverables as well.

* Microsoft Project - create project plan schedule
* Microsoft Word- document project and open other essential documents
* Microsoft Excel - View csv file data and prepare data for cleaning.
* Adobe PDF to edit or create PDF documents
* Jupyter Notebook - Data Science Project Models (Used through Anaconda software), this is also used for data transformation, cleaning and modelling
* Powershell/Command Prompt - Run notebook files and upload to Github

**Cloud-based platforms:** Due to the use of multiple documents and files, file sharing is done easily through collaborative work on cloud based platforms.

* Google docs - Main web application used to document and share documents among the group members for easy collaboration by leaving comments and allowing multiple users to edit simultaneously on the same document
* Google drive - Organised folder shared with all the members containing all the files required for the project
* Google slides - Collaborative presentation slide for the final presentation
* Github- Allows storage for Jupyter Notebook files and other data files.

**Communication platforms:** Working on a project from home means it is hard to keep in touch or update one another, therefore there is great importance in having appropriate communication platforms to easily collaborate with each other.

* Discord: Weekly meetings with just the project group members are held over discord calls and it is also the main form of messaging service for the team.
* Teams: Meeting or catchup session with the sponsor as a group.

**People:** A vital resource for the project to be completed is ultimately the people involved, the sponsor as well as team members.

* Fouad, the Project Sponsor, is responsible for leading the team with an overview of what the project goal is and how it should be approached. This is achieved through various teams meeting to gather as much information and also get relevant feedback for improvement.
* Project Team are responsible for the project planning, analysis, testing and evaluation and meeting the requirements of the sponsor has outlined. Good communication skills between all the members will result in a clear understanding of the project requirements and outcome to be achieved.

## Team Organisation and Structure

The project team follows a team organisational structure that is decentralised and is not segregated into different departments or roles. The reason for this decentralisation is because the individuals in the project team will have a great deal of decision making power once the Project Sponsor provides an overarching direction or plan for the project. Working on a project that is provided by the Sponsor on behalf of BT Financial Group, the Project Sponsor’s main responsibility is to outline the project goal. The managerial instructions given by the Sponsor are to be followed by the Project Officers (Project Team) and a solution is then developed to achieve the project goal.

**Responsibilities:**

Project Sponsor(Fouad)

* Weekly meetings with Project Team on Teams to discuss project progress and provide clarification on any misunderstandings and feedback.
* Prepare project documentation that clearly outlines the information necessary to start the project.
* Review deliverables before submission and provide feedback for improvement before final submission on ilearn.
* Providing other resources that may be helpful for the project's success.
* Answering enquiries on the project that the team might have.

Project Officers/Project Team(Group 6)

* Having regular meetings as a group to update each other of the progress and collect information to assist with the project as a group.
* Meet with Project Sponsor regularly to discuss the progress of the project as well as asking questions to clarify any doubts or confusions.
* Developing a feasibility study and a project plan along with a scoping document.
* Collate data from a trusted source and verify relevancy with the project
* Identify and evaluate the data characteristics to further clean, integrate and format.
* Use various modelling techniques to identify which technique produces the best possible results.
* Make sure to take meeting minutes and clearly document every single part of the project, in case clarification is required at all times.
* Evaluate the results gathered as well as the feedback given to review the process and provide a final project outcome.
* Hand over completed project to Project sponsor

Team Member Roles and Responsibilities

To allocate the roles and responsibilities of each team member of the project, we initially made a list of potential tasks within the project. After creating a list of tasks, based on the team's strengths and weaknesses we were able to decide on each other's responsibilities within the project.

* Ella-
* Keerthana - Algorithm/Model/Method Documentation
* Kurtis- Feature Engineering/ Solution Architecture Documentation
* Nathan-
* Sherry - Model Evaluation Documentation
* Tom - Prototype/MVP/Project Output Documentation

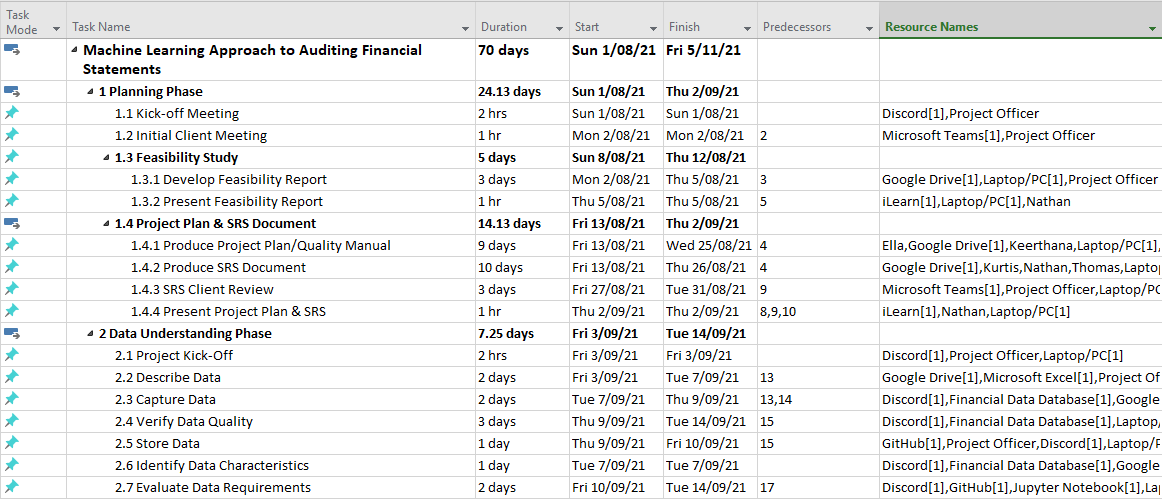
**Organisational Chart**

# 

## 

## Project Schedule

In tackling the project, it is important to condense it to more manageable, tangible tasks that can be accomplished with ease by the project team. The project at hand is composed of six major phases: the Planning, Data Understanding, Data Preparation, Modelling, Evaluation, and Presentation phases. Each phase consists of deliverables that are to be accomplished prior to mark the completion of the current phase and must be presented prior to the commencement of subsequent phases. The deliverables are further broken down into individual units of work called activities that must be completed to fulfill the detailed deliverables (Watt 2014).Below is an outline of all the activities to be performed within each deliverable and phase, alongside its predecessors, and resources required to complete each activity.



## 

## 

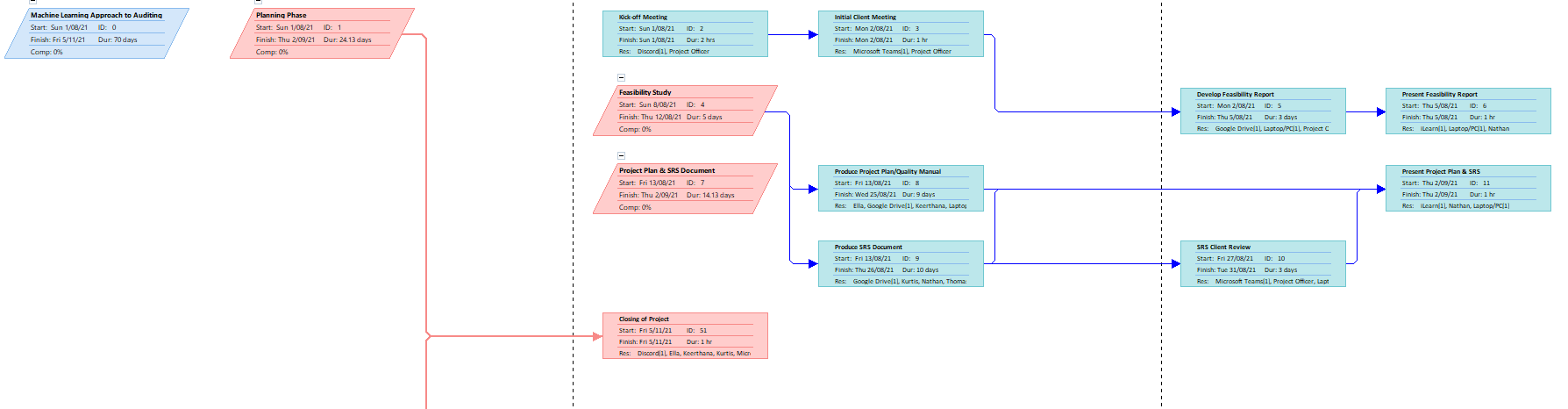
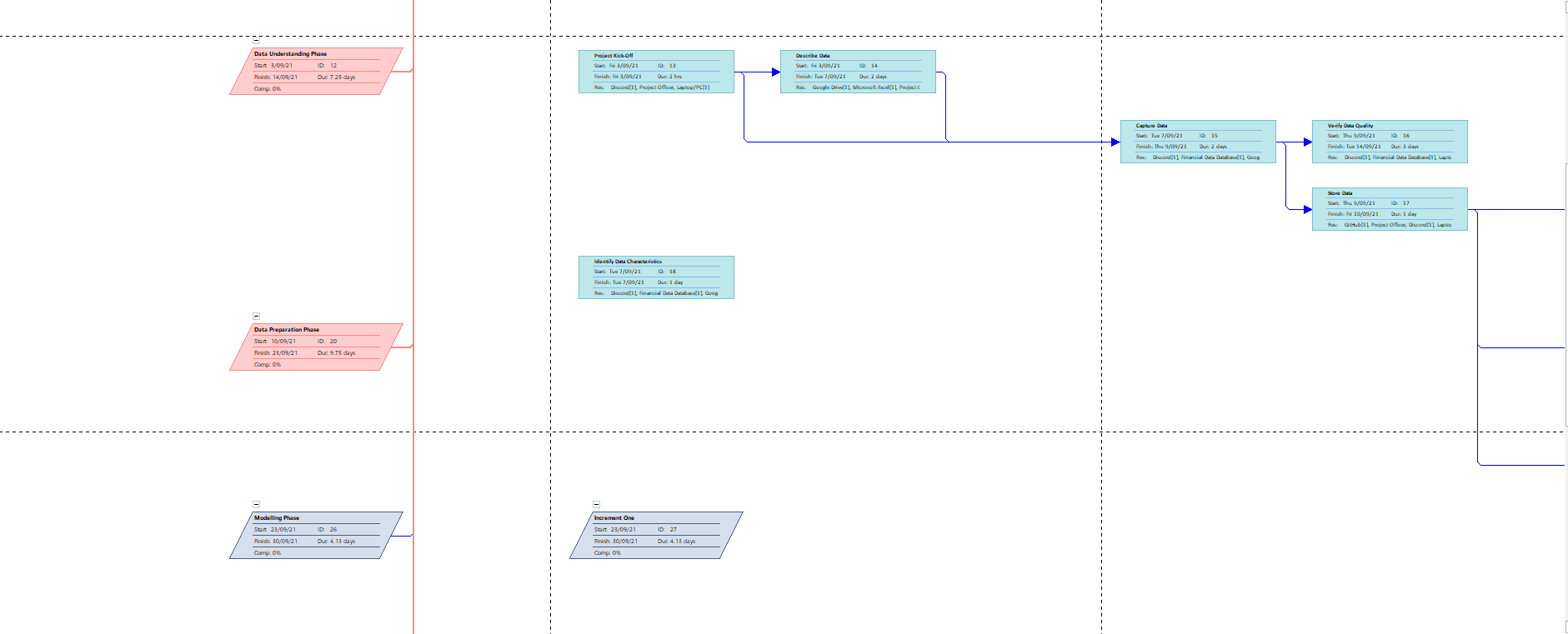
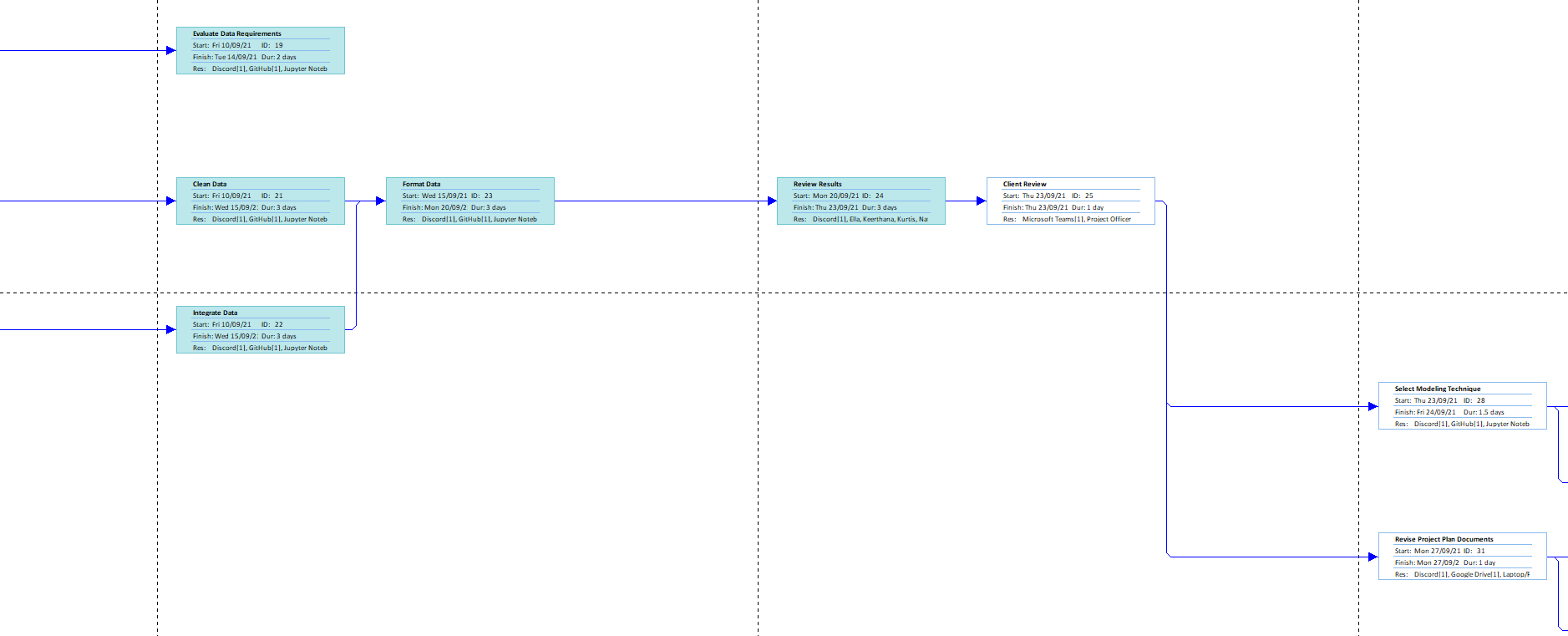
## Timeline

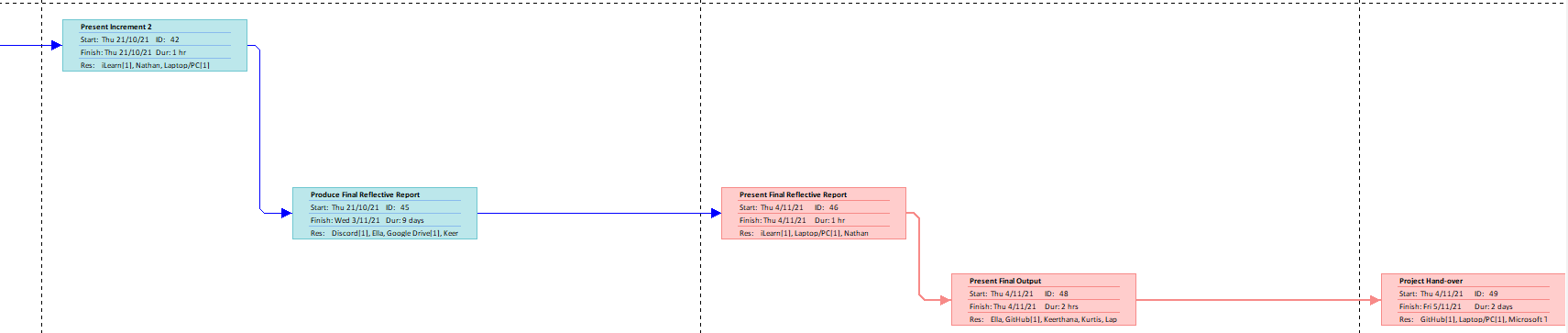
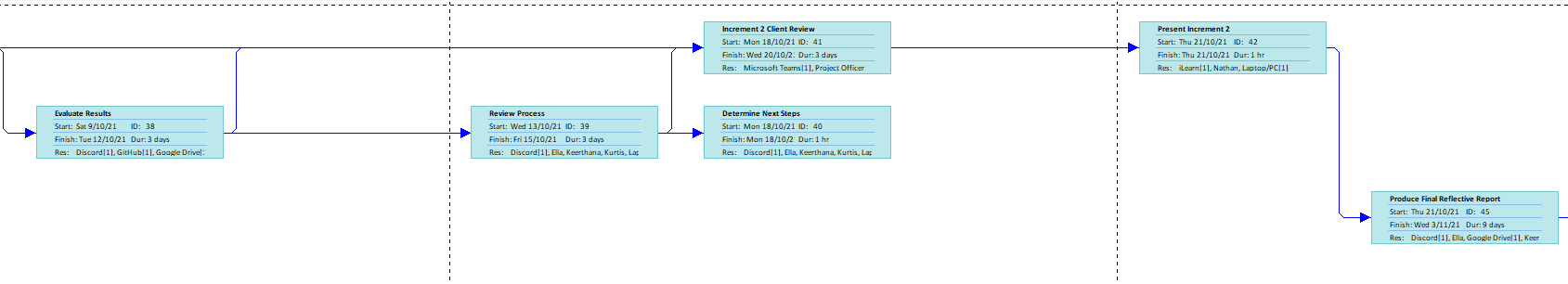
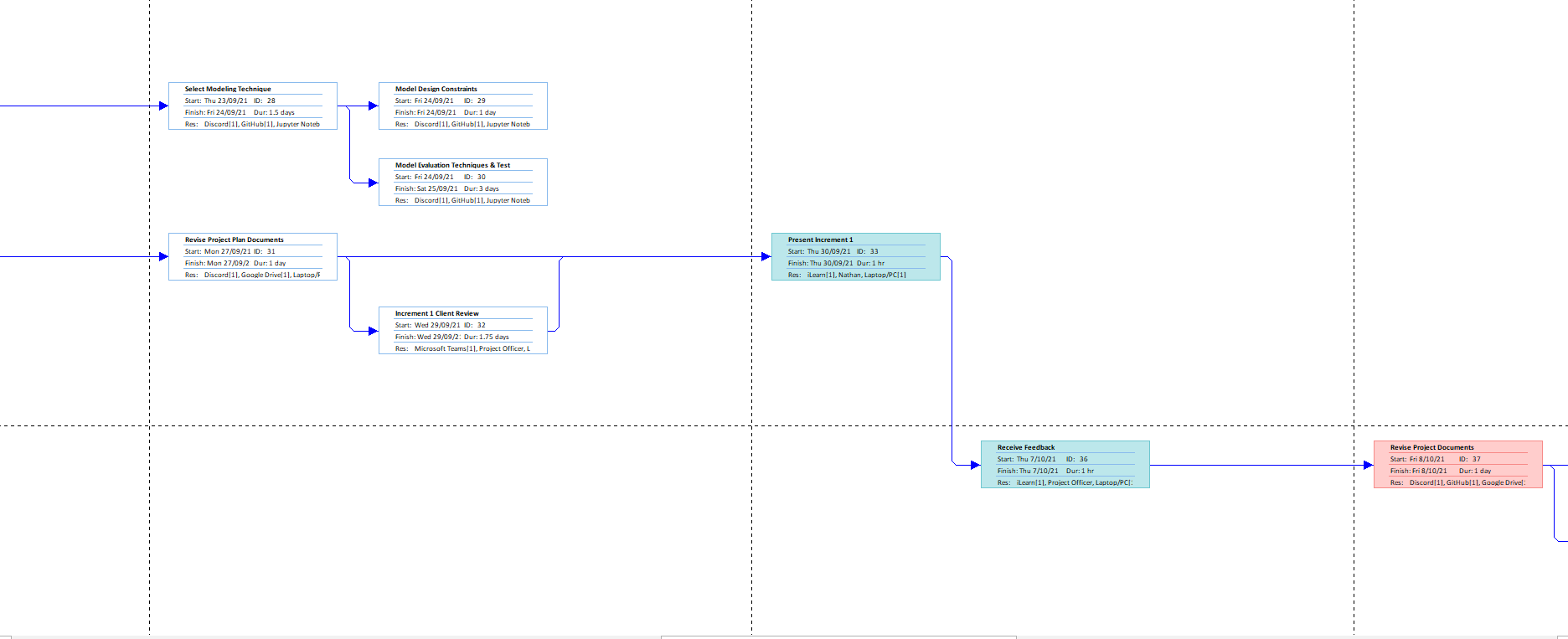
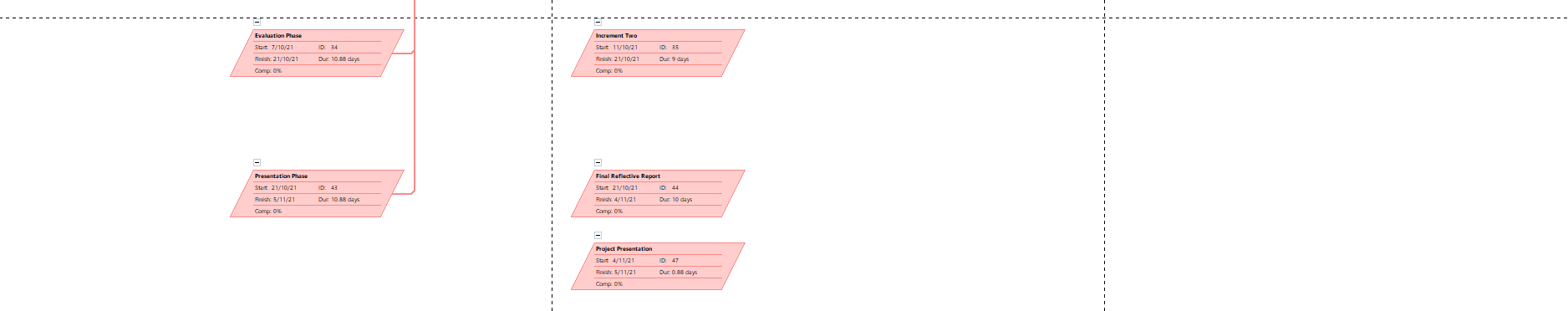
## 

## Resource Sheet

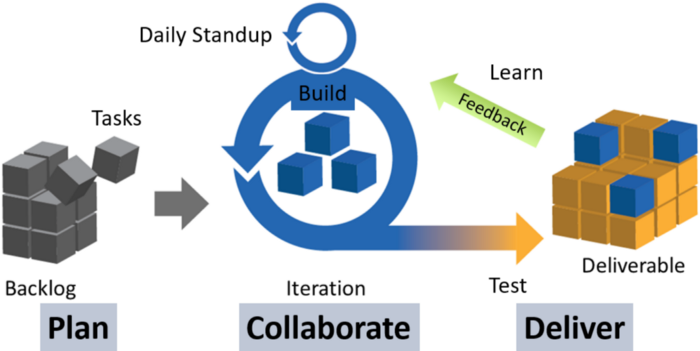
## 

## Network Diagram





## Process model discussed/justified

With the nature and scope of the project at hand, the project team will implement Agile for Data Science approach. Agile Methodology is a strategy to manage a project by breaking it down into smaller, more manageable phases. It incorporates constant stakeholder collaboration and continuous improvement through every working iteration (Wrike 2015). In relation to software and process, is an extension of the Scientific Method that highlights an organised methodology based on hypothesis, observation and learning (Bhatti 2017).

The project team utilises Scrum, an existing framework within the Agile methodology, but has been customised to fit the requirements of the project. There are three specific subsections that comprise the Scrum framework: Roles, Events, and Artifacts.

Scrum Roles consists of a product owner, a scrum master, and the scrum team. Keeping in mind the size of the team and the decision to maintain an equal role and responsibility over the project, the role of the scrum master will not be applicable to this particular life cycle, and all members will be a part of the scrum team.

Furthermore, the project sponsor/client will be responsible for taking on the role as product owner, who will define the potential product, tasks, and objectives of the project.

Scrum determines five main events, all of which are contained in the first event -- the Sprint, which divides the project into a series of mini projects, lasting from one to four weeks (What is Agile Data Science 2019). In one sprint, the following events occur:

**Sprint Planning**

In this event the scrum team determines the Backlog items to tackle, define the deliverables and what is to be delivered by the end of the Sprint phase, and develop a plan of action.

**Weekly Scrum**

This event originally is a daily schedule wherein the team coordinates and comes up with the day’s plan. However, due to the time constraint of the project and the amount of workload to tackle (both on the project and the individual team members’ responsibilities outside of it), a weekly scrum was more befitting to the efficiency of the project team. In this event, the team discusses plans, tasks to be accomplished in the scrum, and updates on activities.

**Sprint Review**

After every Sprint, the team meets with the stakeholders to present the increments and gain insight and/or feedback during the Sprint Review. The Product Owner (or in this case, the project client) is responsible for enlisting what can be done to improve the output in the following phase, and what else is to be done in the next Sprint.

**Sprint Retrospective**

In conclusion of each sprint, the team reflects on improvements that can be made and tangible plans to do so in the next Sprint (What is Agile Data Science 2019).

In utilising an Agile methodology in this project wherein the team will perform strenuous tasks in identifying, collecting, cleaning and understanding data, it will create a process that prioritises not only efficiency in the performance of tasks, agility in accomplishing them, but also providing allowance to make mistakes and adapt to them in a way that is productive to the project’s end product.

The Agile framework appeals to quick action, fail quickly, evaluate and learn, then reattempting with a different approach or improved method (Ng 2020). As stated in the Principles behind the Agile Manifesto (Beedle et al. 2001), “At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly”. Due to the complexity of the project, the team requires opportunities to reevaluate, assess, evolve, and possibly change the plan of action implemented to achieve the end result, and this environment provides just that without having to delay the project timeline.

## Documentation Identified/Discussed

In any project, documentation is a vital aspect to ensuring the success of the final product. It deals with the process of capturing important details and developing documents needed to successfully implement the project. Below are the following documents that have been/will be developed by the team to present comprehensive information.

**Feasibility Report**

A feasibility report presents the team’s assessment of the current issue the client is facing, evaluation of alternative solutions to the problem and concludes which is the most viable for further analysis. With this document, the team was able to not only gain more knowledge into the issue and brainstorm ways in which the issue can be resolved, but also provide ourselves with opportunities to explore various solutions that may not be restricted to one sole solution. This also gave the sponsor a chance to provide feedback on what they believe is the best course of action.

### 

**Project Plan/Quality Manual & Scoping Documents**

The current document presents the finer details of the planning stage of the project lifecycle. This document will serve as a reference point for the team to refer back to throughout the project schedule. The Project Plan involves the management of the project processes while the scoping document specifies the scope of work to be performed by the team to reach the final output.

**Document Revisions**

As the project progresses, there will always be a need to change the plan, this would be a cause of an update or revision of both the project plan, as any changes made would alter the project schedule and possibly the management plans in place; and the SRS document, as direct decisions made in the creation of the document would need to be well documented.

**Final Output**

The team’s main choice of software for this project is Jupyter Notebook. All data will be inputted into Jupyter and stored on the team’s Github repository, which will house both the data and the model itself. This method was selected as it allows a convenient transaction or transfer of larger file sizes over the cloud. Furthermore, all users can access the files on their local machines, work on it and commit any changes to Github, without worrying about other users.

# Quality Manual

## Quality Control and Management

**How will the data be verified? How quality of code be validated?**

Quality Management deals with the process of overseeing all project activities and preserving a determined level of quality, to not only meet written requirements but the clients’ needs and expectations as well. This is done by developing a good relationship with stakeholders to understand what quality means to them (Schwalbe 2018).

A Quality Control and Management plan is created to ensure that every project deliverable conforms to the mentioned level of quality that is acceptable to the expected project output as well as to the stakeholders’ standards. The project team employs the following plan to assess and measure the quality of the project, monitor, and improve it based on the set standards.

**Quality Management Process**

Schwalbe (2018) states that there are three main processes in Project Quality Control management, all of which work towards managing and controlling quality of activities that work towards accomplishing the project: Planning Quality Management, Managing Quality, and Controlling Quality.

**Planning Quality Management**

The initial step deals with identifying quality control requirements and standards relevant to the project, and approach to satisfy them.This information is documented in relation to the listed project objectives, approach, deliverables, and resources available to the team. Any quality activities are documented within this section of the quality manual document, whereas quality control activities relating to project deliverables will be documented on separate documents specific to the scope of its performance.

**Managing Quality Management**

The second step of the process is the actual process of performing the quality management plan, which encompasses the performance of quality control and assurance activities. Quality Control involves evaluating whether the end product meets the quality requirements that are specified for the project, it may also include monitoring and recording results of executing quality activities to assess project performance (Project Management Knowledge 2017), whereas Quality Assurance is the process of auditing quality requirements to ensure appropriate quality standards and operational definitions are used (Project Management Institute 2013).

**Tools for Managing Quality Control**

*Sponsor Feedback from Output*

With the current project, the team requires constant communication with the sponsor to not only understand the project requirements but also receive insight into the project performance. The final output must also meet the client’s needs and expectations, thus the importance of maintaining that relationship to manage the quality of the project deliverable and increments.

*Benchmarking*

This process is used to improve the quality of the project and processes by comparing projects, methods or processes with other comparable organisations (Tran 2017). In utilising benchmarking, the team is able to identify other successful projects, determine the reason it is deemed successful, identify the gaps in quality and performance, then develop measures to improve the project.

**Activities for Managing Quality Assurance**

For a more in-depth explanation of each activity, please refer to the Communications Management section of this document. Quality Assurance activities include the following:

*Project team meetings*

These meetings are held within the team to discuss the weekly progress of the current task being accomplished. This activity gives the team insight into the current state and decides whether there is a need to revise the current plan of action to accommodate any extra resources to a specific task.

*Project review meetings*

This activity occurs after every milestone or prior to the commencement of a following phase in the project life cycle. This meeting allows the team to discuss what was done successfully in the current phase and what can be improved on in the following phase, and make the required changes to the project plan.

*Sponsor meetings*

As the client’s needs and expectations are just as important as accomplishing the project requirements, meeting with them on a weekly basis gives the team the opportunity to identify specifications that the client requires. This time is also used to receive feedback on the output presented to the client; changes which will materialise in the following phase.

**Controlling Quality Management**

The final step of the Quality Management process is control, which monitors and records results of performing the previous activities to assess performance and ensure project outputs are complete, correct, and meet stakeholder expectations (Schwalbe 2018).

*Quality Reviews*

A project quality review is utilised to identify root causes of problems and provide detailed guidance on how to solve them (Stanleigh 2016). This review entails a comprehensive analysis of the entire project to assess all issues and concerns and determine its fundamental cause. The aforementioned assurance activities are an extremely useful and integral aspect of the project analysis. Once this has been performed, a detailed report and recommendations can be presented for further review among the project officers and client to determine the best action to take to achieve overall performance improvement.

*Lessons Learned Reports*

With the information gathered from the quality review, the team members can develop a lessons learned report that details knowledge accrued to the team dealing with issues and obstacles that they were faced with. Documenting and sharing these lessons among the team prevents any repetition of mistakes as well as taking advantage of organisation best practices--innovative approaches and good work practices, can be used to improve future stages of the current project and future projects (Rowe & Sikes 2006).

## Reviews and Audits, Testing, Tools

Data will be collected from a trusted source, and it will be verified to make sure it is relevant to the project (Assumption 9 & 10). After the validity is determined the data characteristics will be identified and evaluated before heading into the cleaning system, where the data will be cleansed of any errors, then transformed into the correct format and finally integrated with the other clean data sets. Then these data sets will be passed through various models, which will determine accuracy and validity of the models.

All of the data gathered and the modelling technique generated will be stored in the GitHub repository so that the team can keep track of the update and all of the work can be reviewed. The work of the project is reviewed by the team in weekly meetings to ensure consistent quality (Assumption 1). The result from generating modelling techniques by Jupyter notebook will be presented to the team to ensure every team member can describe the difference in the model and how to analyse the results. Consequently, all of the modelling techniques used will be analysed and compared to identify which modelling techniques produce the best result. Also, the different data sample such as half-yearly financial data and full-yearly financial data from the same company will be generated by modelling techniques to check the accuracy of the results. To compare the models and algorithms, a confusion matrix is built to indicates the number of test cases correctly and incorrectly categorized. Precision and Recall are evaluated to estimate F1-Score. The project will be reviewed by the client regularly in fortnightly meetings to ensure that client is aware of the progress of this project so that clients can evaluate their satisfaction and dissatisfaction with this project.

## Tracking/Change Management and Tools

There are many ways to have an effective change management process (Assumption 4)

**Identify what will be improved:**

Because most changes aim to improve a technique, a method, or a procedure, it's critical to define the objectives and focus. This also entails identifying project resources as well as various roles and duties that will help to guide the process and lead the project. Most change frameworks realize that identifying what needs to be improved provides a solid foundation for clarity, simplicity, and successful implementation. During the project implementation, it is important to understand the relevancy of data and the difference between modelling techniques.

**Present a plan to the sponsor:**

It's critical to communicate with the sponsor about the solid plan so that the sponsor can provide feedback and make modifications as needed. The sponsors could have a different expectation or point of view than the team do. As a result, for a successful change management process, diverse viewpoints must be considered.

**Usage of data and provide resources:**

Recognizing and allocating resources are crucial aspects of the planning process that encompass the hardware, software, people, and applications that make up the system. Other techniques for re-educating, retraining, and reconsidering goals and actions must also be explored. Many models considered data gathering and analysis as a capability that is underutilized. As a result, regular progress reporting enables for improved communication, effective and timely incentive delivery, and a better measurement of achievements and milestones.

**Communication:**

A comprehensive change management strategy must be defined, prepared, onboarded, and implemented for effective communication. It is essential to communicate on a frequent basis in order to track progress. As a result, wiser judgments can be made about which changes work and which ones do not as the data is being analysed and the modelling techniques are being generated. Furthermore, everyone is aware of which task needs to be modified and the modifications being made.

**Analyse, evaluate, revise, continuously improve**:

Change is a continuous process, and management approaches are frequently altered over a project. To detect and remove barriers, it should be integrated throughout all processes, including communication. Furthermore, continual appraisal and recording of work can aid in ongoing and consistent improvement.

Any modifications to the project's documentation and coding can be tracked via GitHub and Google Docs. Team members can view the source code, leave comments, and edit the repository, ensuring the project’s consistency throughout its life cycle.

**Tools for Quality Management**

The following tools will be used to manage quality:

**GitHub**

The project involved collecting a large amount of data, analysing it with different models and comparing the results to determine which model was most suitable for the project. Therefore, we needed a repository named GitHub to save all the steps of the work (Assumption 9) and to facilitate the discussion and modification of the modelling technique used by the team in order to best meet the requirements of the client. Also, it is important to store the data files gathered from diverse organisations in various industries to maximise the input necessary for the model.

**Google Doc**

Google Doc was used as a platform to document and share resources within the group. It allows multiple group members to leave comments and edit the same document simultaneously. This allows the team to clearly identify the direction and progress of the project work and to discuss it in a timely manner. This important during the COVID-19 pandemic, since the team cannot have face-to-face meeting regularly to discuss about the progress of the project.

**Network Diagram**

A Network Diagram was created and used as a Flow Chart for the project, namely a step by step guide that leads the team to meet sponsor’s requirements of the project. To ensure quality of the project, all of the tasks need to be completed in order. The Network Diagram shows a critical path which gives an indicator that which task is important, if extended, will delay the entire project. This task shown in the critical path must be completed promptly to ensure the quality and timeline of the project.

## Communication

A Communications Management plan aids in ensuring that there is constant flow of communication between members of the team, as well as communication between the project team members and the project stakeholders. The information communicated will help the team and stakeholders make informed decisions on the planning and execution of activities to achieve project success.

This section of the project plan document defines the model of communication required to efficiently and effectively distribute information. The communications plan dictate the following details:

* The approach to communications and need for possible changes
* Frequency in which communication is to be performed
* Distribution of communication: who receives what information, which circles of communication are set, what formats or platforms are used to perform communication
* Escalation processes in place when dealing with communication-related conflict resolution

**Communication Approach**

In order to achieve the communication plan’s objectives, all project team members will take on an active role to not only reach them but also maintain communication on this project.

Any current communication requirements have been documented in the Communications Matrix below. This matrix will serve as a framework for the main modes of communication (what information is communicated, who communicates it, to whom it is communicated to, when it is communicated).

Updates and changes may be required as the project continues to mature. May it be due to internal or external issues directly or indirectly relating to the project. Furthermore, these changes may be required as the situation changes and additional communication is needed to maintain the determined project timeline. As previously stated, all project team members are responsible for preserving communication as well as managing any and all revisions to the Communications plan as the need for it changes. In addition, all approved changes will be documented and updates will be distributed to affected members. Finally, all members will adhere and participate in the communication strategies as stated below.

**Communications Matrix**

The following table identifies the channels of communication organised to ensure that information is disseminated and received by the project team members as well as establishing communication between the team and stakeholders.

| **Communication** | **Format** | **Frequency** | **Distribution** |
| --- | --- | --- | --- |
| Kick-off Meeting | Zoom Meeting | Once | Project team + sponsor + stakeholders |
| Weekly Reports | iLearn forum | Weekly | Project team + stakeholders |
| Sponsor Meetings | Microsoft Teams | Weekly | Project team + sponsor |
| Project Team Meetings | Discord | Weekly | Project team |
| Project Review Meetings | TBD | After deliverable submissions | Project team + sponsor |
| Project Hand-over | TBD | Closing/End of Project | Project team + stakeholders |

**Kick-off Meeting**

* The Kick-off meeting occurs at the beginning of the project lifecycle and its main purpose is to introduce the project team and sponsor, determine the project’s goals, objectives, and deliverables, and review the appropriate approach to reaching these goals. This is to ensure that every member involved is on the same page regarding the work required for completing the project.

**Weekly Reports**

* The Weekly Reports are standard written forms submitted through the iLearn forum to be reviewed by the unit convenor. These reports communicate the team’s progress within the week, the team meetings held or scheduled, tasks to be accomplished within the week and in the near future along with the team member’s responsibilities and task delegation.

**Sponsor Meetings**

* Sponsor Meetings are held weekly through Microsoft Teams with BT Financial’s liaison, Fouad Nagm. This meeting is to update the sponsor on activities, discuss project requirements and seek clarifications, if need be. This correspondence is also extended to emails, if required.

**Project Team meetings**

* The Project Team Meetings are used to discuss the deliverables, condense them to smaller, more manageable tasks, to delegate amongst the team members. This time is also used to update the team on the progress of tasks, information, etc. The meetings are held weekly, on Thursdays through the Discord platform. In addition, further communication is held throughout the week by means of the platform’s messaging features.

**Project Review Meetings**

* After every milestone or deliverable submissions (and prior to tackling the next deliverable of the project), the team will discuss what has been accomplished, gather feedback on what was done correctly, and what areas can be improved on as a team and individually. Moreover, the team presents the project deliverables, consider possible approaches for the next phase of the project lifecycle, and reach a decision on how to tackle it.

**Project Hand-over**

* The Project Hand-over occurs at the closing of the project, wherein the final product and work that has been done over the course of the project lifecycle is submitted to BT Financial. The team will also present the accomplishments achieved in the lifecycle and improvements that can be made in the future.

**Communication Escalation Process**

In circumstances wherein conflicts and disputes regarding project communications occur, it is imperative that it be resolved in a manner that preserves the project schedule, ensure that information communicated is accurate and received timely, in order to prevent difficulties that may arise. Conflict resolution requires a collaborative effort between all project team members to resolve possible disputes with minimal impact to the project.

As such, the team plans on settling communicative issues by reviewing the project (or deliverable) objectives, gathering input from all members, and conferring on the most appropriate approach to reaching the established goals in the most efficient and effective manner.

Regarding conflicts that cannot be resolved amongst the team members and may require external mediation, the team will engage a standard escalation procedure when dealing with these situations. The following table defines the priority levels of the escalation process, the external authorities and the timeframe in which conflict reporting and resolution must be performed to minimise delay.

| **Priority** | **Definition** | **External Authority** | **Timeframe for Reporting & Resolution** |
| --- | --- | --- | --- |
| Level 1 | Major impact to project resulting in a significant impact to project timeline | Unit Convenor, PACE Staff &/or Project Sponsor | Reported as soon as possible, and 5 hours for resolution |
| Level 2 | Medium impact to project resulting in an impact to the project timeline | Unit Convenor | Reported as soon as possible, and 1 day for resolution |

## Conflict Resolution/Negotiation

Since conflict resolution/negotiation is a collaborative endeavour involving all team members while working on this project, it's critical that the team be clear and honest about any difficulties they have or expect to have at any point during the project's development. Problems can develop in a variety of ways, including scheduling and interpretive conflicts, especially with the strain placed on group communication as a result of the COVID-19 lockdown protocols used in New South Wales.

The team needs to negotiate work distribution since team members prefer to spread workloads as equitably as possible depending on the amount of material needed and level of expertise. This guarantees that each member contributes equally to the final result and has equal input, reducing the likelihood of final result dissatisfaction.

In most cases of conflict resolution, the team would communicate with unit convenor or project sponsor first, because project disagreements are generally caused by project interpretation, which is most likely to be resolved by either a unit convenor or a sponsor.

## Standards/Templates/Appendices/Forms

**Templates: Weekly Reports**

| **Weekly Activities:** |
| --- |
| **Progress:** |
| **Meetings:**   * Group Meeting (DD/MM/YY, (TIME), Group call (Discord), Attendees: Keerthana, Nathan, Xinrui(Sherry), Thomas, Ella, Kurtis |
| **Progress Update:** |

# References

Beedle, M, van Bennekum, A, Cockburn, A, Cunningham, W, Fowler, M, Highsmith, J, Hunt, A, Jeffries, R, Kern, J, Marick, B, C. Martin, R, Schwaber, K, Sutherland, J & Thomas, D 2001, *Principles behind the Agile Manifesto*, Agilemanifesto.org, viewed 30 August 2021, <http://agilemanifesto.org/principles.html>.

Bhatti, B 2017, *Agile for Data Science*, Medium, viewed 30 August 2021, <https://medium.com/@thebabar/agile-for-data-science-c8f46daa0dd4>.

Ng, J 2020, *Applying Agile Framework to Data Science Projects*, Medium, viewed 30 August 2021, <https://jnyh.medium.com/applying-agile-methodology-to-data-science-projects-db50ebbef115>.

Project Management Institute 2013, *Perform Quality Assurance in A Guide to the Project Management Body of Knowledge (PMBOK® Guide), Fifth Edition [Book]*, www.oreilly.com, viewed 31 August 2021, <https://www.oreilly.com/library/view/a-guide-to/9781935589679/sub8.2.xhtml>.

Project Management Knowledge 2017, *Control Quality*, Project Management Knowledge, viewed 31 August 2021, <<https://project-management-knowledge.com/definitions/c/control-quality/>>.

Rowe, S & Sikes, S 2006, *Lessons learned: taking it to the next level,* Paper presented at PMI® Global Congress 2006, Seattle, WA, Newtown Square, PA: Project Management Institute.

Schwalbe, K 2018, *Information technology project management*, 9th edn, Cengage, Boston, Ma.

Stanleigh, M 2016, *Project Quality Reviews | Business Improvement Architects*, Business Improvement Architects, viewed 31 August 2021, <https://bia.ca/how-to-undertake-a-project-quality-review-that-will-increase-your-bottom-line/.>.

Tran, L 2017, *Benchmarking in Project Management (2): Internal Benchmarking - InLoox*, www.inloox.com, viewed 31 August 2021, <https://www.inloox.com/company/blog/articles/benchmarking-in-project-management-2-internal-benchmarking/>.

Watt, A 2014, *Project Schedule Planning in Project Management*, Opentextbc.ca, viewed 30 August 2021, <https://opentextbc.ca/projectmanagement/chapter/chapter-10-project-schedule-planning-project-management/>.

What is Agile Data Science 2019, *Scrum for Data Science*, Data Science Project Management, viewed 30 August 2021, <https://www.datascience-pm.com/scrum/>.

Wrike 2019, *What is Agile Methodology in Project Management?*, Wrike.com, viewed 30 August 2021, <https://www.wrike.com/project-management-guide/faq/what-is-agile-methodology-in-project-management/>.

A picture containing drawing, clock

Description automatically generated

***COMP3850 Project Deliverable Certificate***

| Name of Deliverable | *Project Plan/Quality Manual* |
| --- | --- |
| Date Submitted | *DD / MM / YYYY* |
| Project Group Number | *6* |
| Rubric stream being followed for this deliverable (highlight one)  *Note: the feasibility study has the same rubric for all streams.* | *SOFTWARE Rubric*  *GAMES Rubric*  *CYBERSECURITY Rubric*  *DATA SCIENCE Rubric* |

We, the undersigned members of the above Project Group, collectively and individually certify that the above Project Deliverable, as submitted, **is entirely our own work**, other than where explicitly indicated in the deliverable documentation.

| INITIALS | SURNAME | GIVEN NAME | STUDENT NUMBER | SIGNATURE *(IN-PERSON OR DIGITAL)* |
| --- | --- | --- | --- | --- |
| ***KL*** | ***Lazenby*** | ***Kurtis*** | ***45610266*** | ***kl*** |
| ***KK*** | ***Kannan*** | ***Keerthana*** | ***45870969*** | ***kk*** |
| ***NS*** | ***Soares*** | ***Nathan*** | ***45382417*** | ***ns*** |
| ***TB*** | ***Bienasz*** | ***Thomas*** | ***45982201*** | ***tb*** |
| ***ES*** | ***Salafranca*** | ***Ella*** | ***45524742*** | ***es*** |
| ***XC*** | ***Chen*** | ***Xinrui*** | ***44871481*** | ***xc*** |

*© Macquarie University, 2021***List of tasks completed for the deliverable and activities since last deliverable certificate with totals for each individual team member and whole team**

*(copy individual total row for each member and copy pages if more pages needed)*

| Performed by *(Student Names)* | Duration  *(hrs)* | Complexity  *(L, M, H)* | Name of task | Checked by *(Initials)* |
| --- | --- | --- | --- | --- |
| Kurtis Lazenby | 15 | m | Purpose and Context Diagram | kl |
| 15 | m | Data Cleaning, Integration, Formatting | kl |
| **Total** | 30 |  |  |  |
| Nathan Soares | 10 | m | Introduction | ns |
| 10 | m | Data Understanding | ns |
| 10 | m | Modelling | ns |
| **Total** | 30 |  |  |  |
| Thomas Bienasz | 15 | m | Evaluation Requirements and Process | tb |
| 15 | m | Deployment | tb |
| **Total** | 30 |  |  |  |
| **Team Total** |  |  |  |  |

Scoping Document



**Thomas Bienasz**

**Xinrui Chen**

**Keerthana Kannan**

**Kurtis Lazenby**

**Ella Salafranca**

**Nathan Soares**

22-08-2021

COMP3850 - Computing Industry Project

# 

| **Revision Table** | |
| --- | --- |
| *Section:* | *Changes Made:* |
|  |  |
|  |  |
|  |  |
| Deployment - Deliverables | Further explained Deep Audits intentions on how we will deliver our algorithm to our user. Also mentioned new presentation changes that will occur |
| Deployment - Training Support & Documentation | Included more specific instructions on what will be involved in the training process and how it will be carried out |

# 

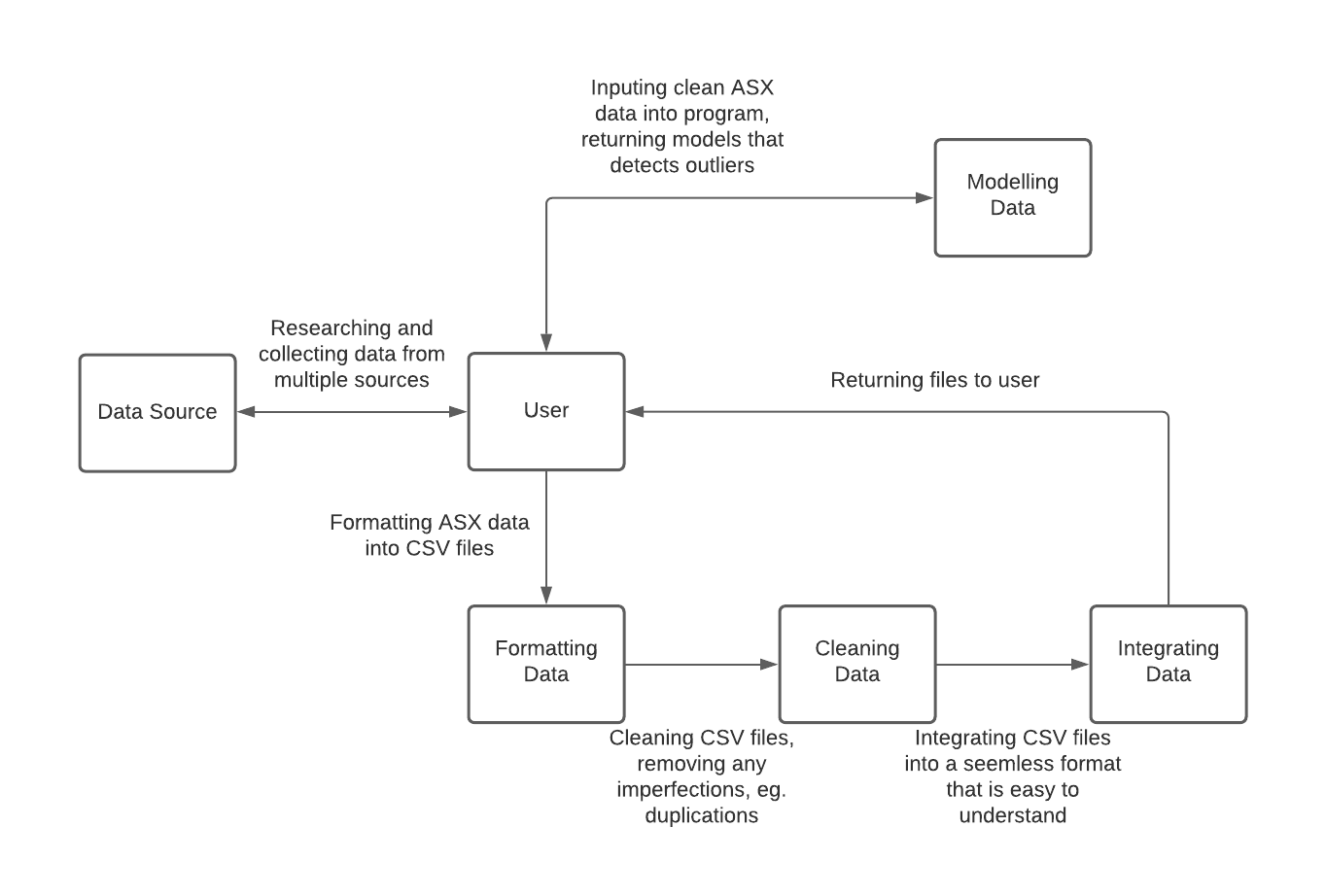
# 

# INTRODUCTION

This document aims to communicate the scope of work for the project, being a machine learning approach to automating the audit of financial statements. The scope of work details the specific objectives and milestones for each phase of the project, as well as technical information regarding the understanding and preparation of the data, proposed modelling techniques, an evaluation process and a deployment plan.

# PURPOSE/CONTEXT DIAGRAM (0.25)

The purpose of this document is to demonstrate the steps we are taking in order to solve the financial auditing process. Below is a context diagram, this outlines how we are transforming ASX financial data into mathematical models in order to identify outliers in the data sets, in turn allowing us to detect suspicious activity occurring in the market.



This context diagram outlines the direction we are taking in this project. It displays how we are going to research, collect, clean and manipulate the data in order to detect financial fraud in the ASX. As seen above, we are researching multiple sources to determine which source has the cleanest data. Once this has been decided, we will pass these datasets throughout our data cleaning process, which will collate all relevant data and place it into an easy to read format, which is then passed back to the user. These clean sets of data will be imported into our python modelling system, which will implement mathematical models to detect outliers in the datasets, which could be a strong indication of financial fraud which may be occurring in the ASX.

# DATA UNDERSTANDING

## Data Description (0.5)

This project relies on financial statement data, which takes the form of income statements and balance sheets which are published on a semi-annual basis. The financial statements should be selected from as many ASX listed companies in as many industries as possible in order to maximise the amount of training data for the model. Since reports are released on a regular basis, a time series component of the data can be captured by comparing reports for a particular firm across each half-yearly reporting period.

## Data Sources & Capturing Data (0.5 + 0.5)

The financial statements for Australian publicly listed firms are freely available on the ASX website within the released reports, alongside all other published announcements, available [here](https://www2.asx.com.au/markets/trade-our-cash-market/announcements.wbc). However, these are provided in a PDF format that includes a vast array of information besides simply the financial statements (e.g., strategic objectives, corporate social responsibility initiatives, etc.). This presents an issue for extracting the desired numerical data from the rest of the report.

As an alternative, Yahoo! Finance provides the isolated income statements and balance sheets of most Australian companies listed on the ASX, however only shows data from 2018 onwards. This may be an issue as we would ideally want to have a longer history of data. As a further preferred alternative, we have requested access to the Company Fundamentals dataset from Sirca, a finance research provider for universities, including Macquarie University. This dataset is extensive and if we are able to gain access, then this will rapidly improve the quality and breadth of the source data we can use for modelling.

Given the timeframe of this project, there will not be any new half-yearly or annual reports released besides those already available on the ASX website, thus there will not be any new financial statements to process. However, if we consider market announcements as supplementary input data to support our model, then we will indeed have new data to capture on a daily basis. For practicality, it would be best to periodically download these releases in bulk as opposed to monitoring them day by day due to the quantity of data.

## Data Storage Environment (0.5)

The data for the project, as well as the model itself, will be stored in the team’s GitHub repository as a central access location. By storing the data in the cloud, the team can ensure that each member is operating on the same datasets. This will also allow the data and the project itself to be easily shared for evaluation purposes by the client, who can download the data and model to their local machine.

## Data Characteristics (0.5)

The quality, structure and completeness of the data is highly dependent on where it is sourced from. This is predominantly due to the variability in how financial statements are presented in reports. While all publicly listed companies on the ASX must abide by the standards set by the Australian Accounting Standards Board, there is still a degree of flexibility in the precise wording and order of reported financial items, as well as if they even appear (e.g. Net Income and Net Profit After Tax are synonymous and either may appear in an income statement). As a result, while the data is relatively well structured in a table format, there will be a processing requirement to ensure that the names of reported items are standardised so they can be effectively compared. The quality of the data is likely to be higher if it has already been pre-processed by sources such as Yahoo! Finance or Sirca, which would alleviate some of the burden on this project in terms of data preparation.

Similarly, some reporting items may appear for some companies and not for others. For example, a bank such as CBA may report provisions for doubtful debts on their income statements with reference to the large volume of loans they have issued. This item is unlikely to appear for a different company such as Woolworths Group which predominantly operates in retail. As a result, the completeness of the data is likely to be an issue as many records will likely contain sector or industry specific reporting items that will be largely absent from other companies. In response, it may make more sense for the model to consider different industries separately because of their different reporting methods.

## Evaluating Data Requirements

The scope of the data collected may go beyond financial statements over the course of the project if it is deemed that valuable insights can be drawn from other elements in the report, e.g. media announcements, notes to the financial statements, cash flow statements or more context-specific numerical reporting items such as acquisitions or share splits. This is dependent on the feasibility of extracting this data from the reports and preparing it into a format suitable for inputting into the model.

# DATA PREPARATION

## Data Cleaning (0.25)

Data cleaning is one of the first steps in our modelling process, as we must remove all duplicate, corrupted and incorrect data. Thus, ensuring that any data that does not belong in the data set is removed. This guarantees that the results from the program are as accurate as they can possibly be as all the contaminated data has been removed.

The majority of the data used in this project will be quarterly and half yearly financial records from all companies listed on the ASX. These reports may be sourced from different locations and will be formatted uniquely for each individual company and industry. Thus, making the cleaning process one of the most important components of our project, as we need the data to be seamless when passing through our program.

In order to achieve these results, we must implement a two-step data cleaning process. Starting with Data analysis, thus being the more crucial step in the process, as we must determine which type of errors and inconsistencies must be removed. This requires a manual inspection of the data before, to validate key pieces of data before it is passed through a cleaning system. This ensures that the quality of the data is high. The second step is a transformation and verification process before the clean data is released back to the user. The transformation process is where the data analysis is implemented, cleaning out any inconsistencies from the files as it passes over the different sets. Then we need to verify whether the transformation has been effective with removing all errors, thus being completed with another manual inspection before preparing it for data integration.

## 

## 

## Data Integration (0.25)

Data integration occurs after all data sets have been properly formatted and cleaned. It involves a process where data sets from multiple sources will be combined into one file, which provides users with a collated version of all the data. This greatly benefits the user as it is an efficient way to collaborate large amounts of data, not only saving the using time but also increasing the accuracy of the data. The data integration method that we will be implementing is the extract, transform and load (ETL Mapping) method. This method copies all of the relevant data from different sources and stores it in a database for transformation, where analysis tools are used in order to integrate the data according to the user specifications.

This project will be using ASX data sets, where our primary sources will be SIRCA and Morning Star. The data is typically represented in a CSV format which makes it easy for integration post cleaning.

## Required Data Formats (0.5)

Data formatting is the last step in the preparation process, where we need to format the data into specific data types, shapes and features. This allows the data to be input into a specific tabular format, where the instances being analyzed in the CSV files align with that of the modelling system.

Our modelling system will be mainly based around Python, which works really well with CSV files when importing data. Most of the financial data sources already provide the data in the format, however if this format is not available we will need to transform this data then convert it into a CSV file without creating any new errors in the data. The data located within these files will mainly be numerical with text for titles. Majority of the non-financial data statements are received in pdf format, at this moment of time this is not an issue as we are manually reviewing the data in these documents. However, if we decide to process these documents for semantic analysis we will need to clean this data as well, and format it accordingly, preferably in PDF files.

# 

# MODELLING

## Modelling Technique Selection (0.5)

We intend to develop an unsupervised clustering model using a technique such as K-means in order to aggregate similar firms based on the items reported in their financial statements. Using this model, the financial statement data of new firms can be compared to the model to determine if there are uncharacteristic outliers in that firm’s reporting items. Based on our approach to utilising the datasets mentioned previously, it is necessary to employ an unsupervised approach as there is a lack of labelled data surrounding fraudulent or suspicious reporting.

Using this approach, the K-means model will distinguish its own categories from the input data and new test data can be compared with these clusters to identify any unusual instances. This will be numerically represented as a probability function. This will be used to give insight into the ultimate question at hand, with the score or probability for a particular firm and the items on its financial statements reflecting how likely that item is to occur, or in other words, whether such a value is unusual or suspicious.

## Model Design Constraints (0.25)

The most significant constraint in designing the model is the structure, quality and completeness of the available training data as well as the rarity and poor definition of instances of the target data (i.e., instances of suspicious numerical reporting).

To address this, we aim to analyse the penalty notices issued by ASIC to ASX listed firms in response to improper disclosure to identify any codifiable rules that may be integrated into the model. This is particularly useful considering that the penalty notices specifically detail what conditions or criteria led to the penalty being imposed. For example, it may be elucidated that a firm’s real EBITDA for a period was materially below the forecasted amount by a certain percentage. This percentage may serve as an indicator of what is considered to be improper reporting.

## Model Evaluation Techniques & Test Data (0.25)

Due to the inconsistent test data mentioned previously, we will likely need to hand-pick training data from firms who have been penalised for improper reporting by ASIC. In terms of test data, we may choose to split the training data into testing and training sets using a technique such as Cross Validation. This is because of the relatively small number of known cases of fraudulent reporting data the team has access to.

In addition, we may integrate known cases of clean reporting data to establish a baseline for the reporting items which are considered typical or non-fraudulent. Selecting these known clean cases may require us to make assumptions regarding the validity of the firm’s reporting, however we can be reasonably confident that large public firms subject to a deal of public scrutiny would be largely free of suspicious reporting. Beyond this, we may consider using the codified rules from the penalty notice analysis to construct a synthetic dataset of known cases of fraudulent reporting and use this as test data for the model to evaluate its ability to operate on new data.

The evaluation of the model’s performance will be directly tied to the ‘score’ it produces which represents the likelihood that a particular firm or reporting item is suspicious. In this sense, the model should exhibit a high degree of confidence on the testing data constructed from known cases of fraud, and conversely should give a low score or probability to known clean cases.

# EVALUATION REQUIREMENTS & PROCESS (0.5)

## Reviewing and Evaluating the Results

The results that we obtain will be reviewed by first creating multiple graphs and charts to make them more readable and clear to us. Once the results are organized into a more readable way, our team can determine if our newly proposed approach is effective or not. This can be determined by comparing our results with the current results and seeing how much better or worse they are. For example you could look at the time it takes for the machine learning approach to spot and determine an anomaly in the data compared to the manual approach which is used today. Similarly, examining the different types of trends that are emerging from the gathered results such as the type of businesses most predominantly found with suspicious acts or what sectors of the financial records are mostly used to manipulate. All of this reviewing of the results will enable our team to examine our analysis in more detail, as a result enable us to establish a well working automated algorithm which will prevent future financial record fraud and more importantly lessen the time frame for identifying questionable behaviour.

## Next Steps to When Outcomes/Results are Examined

The next stages depending on what results have been achieved would be to first of all question ourselves how these results have occurred in this way. Is there some arithmetic that is not getting the results that we expected? Then once we gather our thoughts and discuss it amongst ourselves the team would need to look at how to fix or improve these problems/weaknesses. Also reviewing the results could lead us to further steps like expanding our search criteria of the financial records or even provide more in depth analysis on a particular identified company. Another option would be to compare to other previous groups who attempted this similar project and identify if we are getting similar results as they did.

# DEPLOYMENT (0.5)

## Deliverables

Before implementing our proposed and researched algorithm, we need to be sure that it works on smaller datasets. If problems occur, immediately fix them. It is important to test every aspect of the new algorithm for all sorts of data. Once the testing is completed with our own gathered data, then the machine learning approach can be introduced to look at bigger companies that we did not have previous knowledge of any data manipulation being conducted. The best initial implementation would be to make sure that all our calculations are done to perfection and no errors occur. Moreover, it would be important to think about the delivery to our clients. It's important to ensure the visualisation of all the results that are developed to be easily understandable for our users. Making sure things such as heat maps, line graphs and clear legends are thoroughly created for a proper understanding of what each calculation produces. This mindset would enable us to view the various results received and evaluate further analysis into improving our approach.

## Training, Support & Documentation

When introducing our clients to our new algorithm they would need to be given adequate training in order to understand how to operate it correctly. The training could involve how to identify from the results which businesses have conducted data manipulation in their financial records. In addition, the training would include briefly explaining the specific calculations being done before getting the final results so our clients gain general knowledge of what is happening in the background. If the client has any problems they would be able to reach out to us through a web portal for support service if any problems occur. Also they would be given a phone number so they can reach out to us if any urgent failures or difficulties arise. Furthermore, the client would need to sign and consent to a few documents. One of them is agreeing to the terms and conditions that apply, therefore they will use it in the right way and not for other purposes.

## Monitoring Performance

The main factor to consider is observing how successfully our algorithm has been since the introduction of it. One way to check if the program is performing correctly is to look at the success rate of finding the anomalies within proposed companies. Our main goal is to detect suspicious behaviour of a wide variety of businesses and be certain they are in the wrong. The better our algorithm detects fraudulent behaviour among businesses gives us assurance that our program is doing what it was designed to do. Moreover, the big idea here is to try to create a consistent algorithm which can assure our client and provide confidence that manipulative business will be surely discovered with this new machine learning approach.

## Maintenance

As our newly created machine learning algorithm will most certainly consist of unseen vulnerabilities, our team will therefore provide fast and adequate maintenance. Firstly, after a few days of the full implementation, a general feedback form will be introduced to the clients who use it and asked to fill out. All the reviews that we acquire from the users of our application will be beneficial for us to target the main areas of concern which have been raised. In addition, another maintenance measure would be to perform regular updates and patches so our application can be as secure as possible. If a problem occurs, the client can access our online ‘report a problem’ form which they can send us newly identified errors in our software which would massively help us to maintain the quality of our service.

# CUSTOMER FEEDBACK

|  |  |
| --- | --- |
| Meeting Date & Time | 30/08/2021 at 4:00pm |
| Feedback Received | Overall a well-structured draft that answers the key questions of what is included in the project. Can consider a broader range of data sources and consider possible solutions to the model design constraints |
| Team Response | * Added Morningstar’s DatAnalysis database to possible data sources * Integrated penalty notice approach for codifying possible rules to include within the model |