Decision Tree Analyzer

Applied Project Final Report

By

Guangning Wei

Spring, 2021

A paper submitted in partial fulfillment of the requirements for the degree of

Master of Science in Management and Systems

at the

Division of Programs in Business

School of Professional Studies

New York University

Table of Contents

Declaration

I, Guangning Wei, declare that this project report submitted by me to the School of Professional Studies, New York University, in partial fulfillment of the requirement for the award of the degree of Master of Science in Management and Systems, is a record of project work carried out by me under the guidance of Dr. Andres Fortino, NYU Clinical Assistant Professor of Management and Systems. I grant powers of discretion to the Division of Programs in Business, School of Professional Studies, and New York University to allow this report to be copied in part or whole without further reference to me. The permission covers only copies made for study purposes or inclusion in the Division of Programs in Business, School of Professional Studies, and New York University research publications, subject to normal acknowledgment conditions. I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, to award any other degree or diploma in this institute or any other institute or university.

Acknowledgments

I sincerely thank Dr. Andres Fortino for his contribution as the sponsor of this project and as a mentor. I also want to thank all the instructors in the Management and Systems program I have taken courses with and learned a great deal. I would also like to thank the developers from the MTU module where their codes provided clear guidance for this project.

Abstract

This project aims to use R language to develop a decision-tree module that is working correctly under software Jamovi. This project will create an easy teaching tool for the NYU MASY program, allowing faculties and students to do decision-tree analysis without typing R codes by themselves. Also, they may just import the data into Jamovi and use the decision-tree module to perform analysis.

This module will be developed by following the Jamovi module creation tutorials; all the R codes will be written in R studio within the requirement of the Jamovi environment (jmvcore, jmvtools). The required R packages knowledge of decision-tree module development are R6, caret, rpart, class, and rpart.plot. These R packages were mainly used when developing the decision-tree module

The client of this project will have full access to the Jamovi module from the GitHub repository (link:https://github.com/guangningwei?tab=repositories). The descriptions on GitHub will guide the users on installing and using the module on both Windows-based PCs and Apple Macs to eliminate the difficulties of finding an appropriate platform. On the other hand, this module could be used by those who may use it for research purposes. At last, the project is cost-effective with no extra costs.

Abbreviations and Definitions

R: A programming language and free software environment for statistical computing and graphics

R studio: RStudio is an integrated development environment for R, a programming language for statistical computing and graphics.

Jmvcore: A framework for creating rich interactive analyses for the jamovi platform

Jmvtools: An R package, analogous to devtools, makes it easy to develop jamovi modules.

rpart: Recursive partitioning for classification, regression, and survival trees.

rpart.plot: Plot 'rpart' models. Extends plot.rpart() and text.rpart() in the 'rpart' package.

Decision tree analysis: A Decision Tree Analysis is a graphic representation of various alternative solutions to solve a problem. The manner of illustrating proves to be decisive when making a choice

Introduction

Company & Background Information

New York University (NYU) is a private research university founded in 1831 and located in New York City. NYU has many campuses and academic centers worldwide, such as Stern School of Business and Professional Colleges.

The Office of Admissions of NYU School of Professional Studies and the Management and Systems program (MASY) is at 7 East 12th Street, Suite 921, NY, NY 10003. This course aims to provide students with experiential learning opportunities to help them develop and strengthen management and leadership skills and gain relevant Comprehensive knowledge of current information technology. The NYU MASY program provides many competitive courses and learning opportunities for students in different fields of study. To help students learn more about data analysis skills, NYU MASY prepared courses like data mining, data analysis, decision tree analysis, and so on.

Sponsor Information

Dr. Andres Fortino - Clinical Associate Professor of Management and Systems.
He is the sponsor of this project and will supervise the project's progress in three months, and finally receive and validate the project that meets the requirements.

Problem Description/Opportunity

NYU MASY offered many data science courses to teach students to become experts in this industry. Professors would like to use some software to improve teaching quality and help students better understand the course materials.

The client is teaching data analysis. He used software to help students better understand the concepts; however, most of the software in the market is not free, and somehow the cost is way much higher than the benefit of purchasing it. The client wishes to teach decision trees by the end of each semester, and he finds Jamovi that is free and has decision tree analysis. It is accessible to the public and welcome developers to develop various analysis by themselves. Under the Jamovi environment, a decision-tree module was built and programmed with R language before by the Michigan Technological University (MTU); however, it is not working correctly on the client computers. Firstly, this project has to solve the functionality of the module instead of trying to find the issues of the module created by the MTU. The better solution would be recreating another one by following the tutorials of creating analysis in Jamovi. The second problem is how to fix the issue of running platforms. Nowadays, people mostly used Windows or Mac; this project has to develop the decision tree module that working on both systems correctly. At last, the module should be able to free to the public, and it can be accomplished by using GitHub.

The project's final deliverables must be able to meet the requirements of the client, which are: easy to use as a teaching tool, and tool for researching purpose. It also required an installation guide and user manual for the module provided on the final delivery day.

Importance of the project

The decision tree is a predictive model in machine learning; it represents a mapping relationship between object attributes and object values. Each node in the tree represents an object, and each bifurcation path represents a possible attribute value, and each leaf node corresponds to the value of the object represented by the path from the root node to the leaf node. The decision tree has only a single output. If you want to have a complex output, you can build an independent decision tree to handle different outputs. The decision tree is a frequently used technique in data mining, which can be used to analyze data, and it can also be used to make predictions.

The client in NYU MASY is teaching data analysis-related courses to best express the decision tree concepts. He requires an accurate decision tree module that can run in Jamovi to improve teaching quality.

Alternate Solutions Evaluated

There are solutions for the client to solve the issues; however, creating a new module has been select due to the efficiency and effectiveness after careful consideration and discussions with the clients.

1. The first alternative solution would be to let the students purchase the software that contains decision tree analysis. (A)

   Pros: By doing so, students can learn how to do the data analytics work without knowing too much R programing language. It is easy to teach and work with instructors and students.

   Cons: This will cost students money to meet the teaching purpose for most students. It has cost over benefits by just subscribe to one-month software to implement decision-tree analysis and never use again. It can't be guaranteed that software can work on both Mac and Windows systems.

2. The second alternative solution would be teaching students R code for decision tree analysis. (B)

   Pros: Students can understand the decision tree concept much better by typing in the code. It also free because R studio and R packages are free to download.

   Cons: Students have various R programming skills; it is hard to improve the course efficiency. On the other hand, Windows and Mac systems are different when performing coding analysis, and it will increase the difficulties for students to understand the decision tree concept.

3. The third alternative solution would be just developing a decision module that working on Jamovi for both Macs and Windows. (C)

Pros: Client and students can download and use it free with clear guidance. Also, because the module is designed specifically to the requirements of the client. It will serve the purposes effectively.

Cons: The module's further development may be restricted because it severs the requirements only. Client and students may question the accuracy of the module because some R language experts didn’t develop it.

Solution Evaluation Criteria

Developing Evaluation Criteria

If the solution meets NYU MASY's exception on quality.

If the solution meets NYU MASY's mid to long-term projections.

Cost Related Evaluation Criteria

If the solution meets NYU MASY's requirement on cost.

Using Evaluation Criteria

If the solution meets NYU MASY's objectives of providing an easy-to-use tool.

If the solution meets NYU MASY's requirement on teaching materials.

Selection Rational

The following criteria were established to evaluate solutions

The weights for each criterion:

Easy to use	Cost	Developing difficulties	Easy to download	Further development
20%	50%	10%	15%	5%

They are ranking from 0-5, from least to most respectively.

	Easy to use	Cost	Developing difficulties	Easy to download	Further development	Expectation score
A	4	5	1	3	2	3.95
B	1	2	3	5	4	2.45
C	4	1	2	5	3	2.3

In the selection phase, we had various meetings with the client and came to an agreement on the specification requirements of this project. After meeting with the client, the project team found five criteria that can affect the selection of solutions. To select the most appropriate solution, we had used a weighted average approach designed weights to evaluate the expectation score for three alternatives.

The cost is the most crucial factor to the client because it only uses as a decision tree analysis tool. Avoiding unnecessary expenses is extremely important. The second most weighted factor is easy to use. We have to meet the requirement of an easy-using tool for teaching purposes. The next one is easy to download, where it relates to the provision of free download and accessibilities for everyone. The developing difficulties are on the developer's view of the project. Can the developer finish the development on time? Does the developer is skilled in completing the job? That is also another factor we want to emphasize. The last criterion is further development. The first two solutions don’t require any further action or updates because it uses a paid application or just using R codes. In the third solution, we had tried to find the best

way to solve the problems of facing updates on Jamovi by locking down the version of the software.

In the overall score listed in the above table, the lower the score, the better solution; therefore, we are selecting the third alternative as the project solution.

Approach and Methodology

This project used Porter's five forces analysis, stakeholders' analysis, WBS, and other analysis to accomplish a waterfall approach. The project will have five different stages, which are initiation, planning, execution, control, and closeout.

The initiation phase is the project's beginning stage. This is the start of the project, and the goal of this phase is to define the project at a broad level. In order to get the project start, the project team is preparing all the documents and deliver it to project sponsor to sign on it at this stage. After all the terms agreed with the project sponsor, the project team can move on to the next step.

The second stage is planning phase which the project team will work on the project charter and WBS. This phase is key to successful project management and focus on developing a project plan that the project team must follow. During this phase, project manager needs to do the literature review, as well as work breakdown structure, risk management plan, project plan. Beyond all of that, the project manager must build consistent communication channel with project client, thus to discuss the requirements of the project.

In the third stage which is the execution stage, in this phase, the project team will follow the project charter determined in the previous step. Also at this stage, the project team will test the module back and forth multiple times to make sure it meets the requirements asked by the client.

The fourth stage is controlling stage where the project team was required to documentation all the process in order to compare monitor the project progress. During this stage the project team will continually track the performance of the current project progress. If there

were possible risks that may affect the project's progress, it should make change by submitting change management plan.

Closeout: Those processes performed to formally complete or close of the project. Once the project is complete, the project team will need to prepare the final project report as an official completion of the project for the capstone courses. Also, the final presentation must be prepared as well.

Level	WBS Code	Element Name	Definition	Due By
1	1	Widget Management System	All work to implement a new widget management system.	
2	1.1	Initiation	The work to initiate the project.	
3	1.1.1	Project Determination	Decide which project to take	January 10th 2021
3	1.1.2	Develop Project Proposal	Project Manager to develop the proposal with deliverables.	January 18th 2021
3	1.1.3	Submit Project Proposal	Delivery the proposal to the project sponsor for approve.	February 1st 2021
3	1.1.4	Project Sponsor Agreement and Sponsor Acceptance	Project Manager develops the Project Sponsor Agreement and Sponsor Acceptance	February 25th 2021
3	1.1.5	Project Sponsor Agreement and Sponsor Acceptance Signed/Approved	The Project Sponsor signs Project Sponsor Agreement and Sponsor Acceptance which authorizes the Project Manager to move to the Planning Process.	February 25th 2021
2	1.2	Planning	The work for the planning process for the project.	
3	1.2.1	Project Kickoff Meeting Time Setup	Project Manager set up a time to meet with Project Sponsor and decide weekly meeting time.	February 1st 2021
3	1.2.2	Determine the requirements of the project	The Project Manager determines the project required resource	February 1st 2021
3	1.2.3	Project Kickoff Meeting	The planning process is officially started with a project kickoff meeting which includes the Project Manager and Project Sponsor	February 4th 2021
3	1.2.4	Develop Project Plan	Project Manager the team develops the project plan.	February 5th 2021
3	1.2.5	Submit Project Plan	Project Manager submits the project plan for approval.	February 6th 2021

3	1.2.6	Milestone: Project Plan Approval by Meeting with Project Sponsor	The project plan is discuses between and the Project Manager and Project Sponsor, then the Project manager has permission to proceed to execute the project according to the project plan.	February 11th 2021
2	1.3	Execution	Work involved to execute the project.	
3	1.3.1	Project Meeting with Specification Requirements	The planning process is officially started with a consent on the specification requirements in the meeting between Project Sponsor and the Project Manager	February 11th 2021
3	1.3.2	Install developing software and packages	The project is start with the required software and packages.	February 12th 2021
3	1.3.3	Develop System selection (Mac OS, Linux, Windows)	The project manager decides the developing system for the project between MacOS, Linux Windows.	February 12th 2021
3	1.3.4	Develop System Alternatives If Possible	The project manager decides the alternative developing system for the project between rest of the two after the module works on the previous system working correctly.	April 30th 2021
3	1.3.5	Draft Module Finished	The draft module finished.	April 20th 2021
3	1.3.6	Testing Phase	The module is tested with sample data.	May 1st 2021
3	1.3.7	Install the Module on Both Systems	The final module is installed on application for both systems.	May 2nd 2021
3	1.3.8	Video Made for Introduction of the Module	Create a video for demonstration of how to use this module	May 5th 2021
3	1.3.9	Upload to the public as an open source.	Modules upload to the public website and prepare for downloading	May 3rd 2021

2	1.4	Control	The work involved for the control process of the project.	
3	1.4.1	Project Management	Overall project management for the project.	
3	1.4.2	Project Status Meetings	Weekly meetings with reports on the progress of the project.	Every Thursday at 11 am for the meeting if needed.
3	1.4.3	Update Project Management Plan	Project Manager updates the Project Management Plan as the project progresses.	Every Wednesday based on the progress or changes for the project.
2	1.5	Closeout	The work to close-out the project.	
3	1.5.1	Submit Sponsor Acceptance to Sign the Last Part	Project Sponsor signed the last part to agree the accomplishment of the project.	May 10th 2021
3	1.5.2	Update Files/Weekly Summary	All files should be regulated and integrated all the weekly summary.	May 8th 2021
3	1.5.3	Presentation of the Project and Gain formal Acceptance	Giving the presentation of the project and illustrate the module on both systems.	May 10th 2021
3	1.5.4	Upload to GitHub Repository	The files of the module must upload to GitHub repository as open source	May 9th 2021
3	1.5.5	Develop the Installation guide and User manual	Having installation guide and user manual ready to turn in with the module.	May 9th 2021

Project Objectives and Metrics

Goal of the project

The project's goal is to create a computer-based tool to perform decision tree analysis in the Jamovi R interface program. The module must work on both Mac and Windows versions of Jamovi, requiring two different repositories for each system. The repositories at GitHub must provide installation guidance and a user manual to describe how to use the module properly on both systems. Also, the module must implement decision tree analysis smoothly without any error messages and becomes the teaching tool in the class.

Project Deliverables and Metrics

- Objective 1 – Convert the discussions with the client into well-defined specification requirements, including all the features and functions for this module.

Measurement: Delivery the typed document in mid-February and go back forth with the client to discuss any changes.

- Objective 2 - Build a Jamovi add-on to perform decision tree analysis for numerical and categorical target variables.

Measurement: Delivery in R language app by the end of the semester.

- Objective 3 - Creating a Git Hub repository of the final project files and deliverables.

Measurement: Create the link for Git Hub with all the module files and deliverables to make it publicly available for anyone to use for decision-tree analysis. Finally, hand in

the link by the end of the semester. https://github.com/guangningwei?tab=repositories

- Objective 4 - Deliver a ready-to-install module to Jamovi with clear user instructions for installation and use.

Measurement: Delivering the module, installation guide, and user manual by the end of the semester.

- Objective 5 - Test the validity of the algorithm using client-defined data sets.

Measurement: The final product will be delivered by the end of the semester with permission to use by showing the functionalities in the presentation.

Risk Analysis

Risk Management is always required for project management. The project team can identify, assess, evaluate, treat, and monitor the risks through the project cycle. Within this project, risks can come from various dimensions, including software, operation, external events. The project team first defines all possible risks, then the project team is doing brainstorming on a risk matrix based on all defined risks and assigns the probability of occurrence of each risk.

Probability of occurrence:

Expected: High possibility of occurring.

Possible: The less possibility of occurring.

Very Unlikely: The low possibility of occurring.

Exposure:

High: high severity impact on the project.

Moderate: Some impact on the project.

Slight: Minor or no impact on the project

	RISK (exposure)			
Probability (of occurrence)		1.Slight	2. Moderate	3. High
	1. Very Unlikely	4		
	2. Possible		3	
	3. Expected	2	1	

After analysis of matrix risks, we found that Risk 1 requires a contingency plan due to it

falls in the red section.

Risk	Description	Probability (1-3)	Exposure (1-3)	Contingency Plan
1	Jamovi has different capability on mac and windows	3	2	Ensure the client is aware that the module may run on different versions of Jamovi on mac and windows.
2	The developer may lack the knowledge of programming in R because he is not familiar with R	3	1	
3	Mac probably is not an excellent system to develop the module	2	2	
4	The project manager is lacking the knowledge of creating a repository on GitHub for the module	1	1	

Issues Encountered

While working on the project, the team encountered some issues. All of the problems the team faced are minor issues which do not have major impact on project. All issues were solved immediately once indicated so that the project was able to finish on time with high quality. Here is all type of issues project team faced in the duration of the project.

The first issue the team faced was using command Jmvcore and command Jmvtools to create a module in RStudio. Unlike develop an R language packages, lot parameters have to be used to create a Jamovi package. In order to pursue further module development, the developer had to go through the tutorials of how to create a module with
t-test analysis.

The second issue appeared when the developer tried to create the module without add any analysis inside the module. Due to the specialties of Jamovi, the module must install before initiating any further analysis. Unfortunately, the developer has a MacBook with Big Sur system; the Jamovi module can't install properly in mac OS. So, the developer switched to the Windows PC and solved the issues.

The third issue stated when the developer started implementing a decision tree analysis in the module. There are lots of parameters showed within the module that was created. It has to be changed into the parameters for decision tree formulas. The transformation was odd in Jamovi, so the developer studied from the MTU module and then applied and updated it to its module.

Project Chronology and Critique

The tasks and duration of the project are shown below:

Level	Element Name	Start By	Due By	Durations (days)
1	Project Determination	01/01/2021	01/10/2021	10
2	Project Proposal	01/10/2021	01/18/2021	8
3	Literature Review	02/04/2021	02/24/2021	1
4	Situational Analysis and Cost Benefit Analysis	02/24/2021	03/11/2021	17
5	Work Break Down Schedule	02/24/2021	03/01/2021	17
6	Sponsor Agreement and Project Sponsor Acceptance	03/01/2021	03/08/2021	7
7	Project Charter	03/08/2021	03/15/2021	7
8	Milestone: Project Plan Approval	03/15/2021	03/22/2021	7
9	Status Report A	03/22/2021	03/29/2021	7
10	Risk Management	03/29/2021	04/12/2021	13
11	Status Report B	03/29/2021	04/12/2021	13
12	Project Management	04/12/2021	04/15/2021	3
13	Project Sponsor Acceptance- Final Project Signoff	05/01/2021	05/03/2021	2
14	Update Files/ Documents	04/15/2021	04/25/2021	10
15	Update Project Management	04/25/2021	05/02/2021	7
16	Final Presentation	05/03/2021	05/10/2021	7
17	Final Report	05/03/2021	05/10/2021	7

The entire project was running smoothly; however, the project management part still needs updating much frequently. This is because the developing progress is faster than expectations, there are short of time to update the management plan frequently.

Lessons Learned

The whole project was able deliver as planed with expected quality and in time and this could not have been done without contribution and help from all team members and sponsor. During the whole project implementation, team member has learned in the following criteria:

Jamovi software:

1. Not every software can used same version on both Mac and Windows.
2. Each software has its own way of implementing R language, it cost time to transformed from one to another for other codes developments.

Project Management:

1. Think more than what was expected at the beginning.
2. Kickoff meeting with client and going back and forth to generate a requirement specification.
3. Keep the client and stakeholders who involve in the project updated with newest project's progress.
4. Control the communication quality and frequency.
5. Prepare the contingency plan for the possible risks occur.

Analytical skills:

1. Knowledge of decision tree and random forest.
2. Data visualization is important for any data analysis progress.

Conclusion and Summary

The project should satisfy the project client's expectations, the graduate school that offers this module development to make the course much exciting and practical than other graduate schools. In this case, it requires that this module should actively use instead of having the project client looking for a substitution.

The project sponsor is running connections with the project manager; the project team essentially keeping client updated on the project's progress and negotiate the possible changes add to the project.

Monitoring is a good approach for having the project's module more competitive in the market. Adding the efficient features from other's modules and software and make it fit with the client's expectations is essential. Also, many GitHub developers had experience developing such tools, and they may offer some ideas for improvements of the project. And that is also why this module is free to download for everyone.

The final module finished in this project would be used mainly by the students who are in the NYU SPS program, as the project manager, it should be responsible for listening carefully about their thoughts on the final product if possible. Actually, the project was finished ahead of the project plan, and the client use it as a teaching tool in the class and received good comments.

Limitations, Recommendations and Scope for Future Work

Even this project was able to deliver as excepted, there are still some limitations within this project and some of the limitations may be improved in the future similar projects in NYU MASY.

First of all, this project is only focus on the decision tree model building, it has the problem of causing overfitting the test and train data set. And also because of the random forest analysis is not a deliverable of this project. It will definitely have the issue of overfitting, and we can't overcome that issue now.

Secondly, Jamovi wasn't a good public software that trusted by Apple, it always has some running issues when actually developing the module or implement analysis. This is not the fault of Jamovi itself, we can easily see that Jamovi has some limitations on Macs. In the future work for the project managers, they may have to spend more time to study on which version of Jamovi can run on their systems. They can't learn from the previous projects because there is no experience can learn caused each module served different purpose of data analysis.

The third one is a limitation that would be the further development works. It because of the purpose of this project is to let students can easily use decision tree analysis on both Mac and Windows. Therefore, the developer developed the module on Mac is only running correctly when the Jamovi version is 1.8.0. on the other hand, the same module built for Windows is only running properly when the Jamovi version is 1.2.27. If the future students or other project manager trying to add more analysis, it may have faced some problems on integration.

Literature Survey

1. Introduction

1.1 Summary

Rapid growth for technology in the recent decades changed the way for decision-makers to implement their decision-making process. Also, in the age of big data, data mining becomes much smoother and has brought a lot of convenience for decision-makers who rely heavily on data set (Yang, 2013). Under these significant changes, the decision-tree is becoming more flexible than before. Due to the software development, tools utilized, and algorithm evolved, decision-makers can easily add features to their decision-tree models for multiple purposes. This paper will provide insights into decision-tree models and the decision-making process, associated with ten peer-viewed journals and scholarly articles. This paper will then conclude how it associates with the current project of making a decision-tree model for the client.

1.2 Decision-tree

The reason to create a decision-tree model in real-life to assist the decision-making process because people are irrational (Kamiński, 2018). Usually, the decision-makers assess the probability of success on its actions and only see there are success and failure, which is 50% for each; however, due to non-stochastic distributional uncertainty, there is no single expected value that can made by only applying this sense (Kamiński, 2018). The decision-tree model would help decision-makers implement a better decision-making process after more to concern the random distribution, then they may compare all the alternative decisions, instead of just measuring each decision either 50% chance to success or 50% chance to fail.

The decision tree is an exciting module that may build easily for decision-makers; the reason to call it an interesting one is because every decision-tree model may have same functionalities on a particular decision-making process, and due to the algorithm difference, they

may differ at other decision-making processes (Jena, 2020). For example, there are two decision-tree models developed with a standard algorithm, and another one was created in another more sophisticated algorithm. Then let them both generate a tree of the data set. There is uneasy to identify two decision-tree models that one better; however if decision-makers change the amount of data input, there might be some difference. This is because the way they are designed is for various purposes, but when put into an environment of having the same decision-making capabilities, they will give similar results, but if changed the environment, there might be one model that exceeds its ability.

In developing decision-tree models, the developers would always have to understand the decision tree's structure. Beyond that, they make changes by adding various features into the models to change the decision-tree models' performance (Kamiński, 2018). There are two approaches for developing better decision-tree models. The first approach believed that making selections of attributes and then placed into a decision tree at appropriate positions is extremely important, such as adding measurement as information gain, Gini index, and diversity index (Shaheen, 2020). The benefit of adding these attributes into the decision-tree model is to let the model becomes more data sensitive and returns much more accurate results. The second approach is to integrate the data mining process with machine learning, thus making the decision tree models extract data set much more accessible and have the ability to eliminate the unusual data or replace it with representative data (Yang, 2013). Both approaches are tremendous but depend on developers' preference. The former required expert knowledge on the attributes, and the latter needed a better understanding of machine learning for the data mining process.

Visualization of the outputs could help understand the essentials of results, which is also an essential feature for the current decision-tree modules developing process (Liu, 2007). Indeed,

today's software could help us select the best outcome once the process is finished; however, instead of having only specific "best" outcomes, as a decision-maker. They would like to see other outcomes for comparison. It is because the best results come out from the computers may not be the best selection for decisions made (Liu, 2007). A quick example that happens in the stock market to understand the importance of having visualization. A day trader can either take a long position or short position for an option of a company to assume whether the stock price is going up or down. Suppose that the trader is using the probability and volatility as attributes in his decision tree analyzer where calculate the future movement of an option. The analyzer used inputs and machine learning on the stock price's past trends, then giving positive returns of 'long the option' that trader should make today. It will not show other alternatives for comparison, which are 'negative returns of long the option,' 'positive returns of short the option,' and 'negative returns of short the option.' So there comes the importance of having visualization for the decision tree models. The decision-maker can then compare the results and integrate them with market news to select the best decision.

The decision-tree model also has its biases. First is selection biases and majority happened while the decision-makers are selecting variables before inputting them into the tree (Kam, 2007). This also could be viewed as granting wrong weights for each variable as well. So, as decision-makers, we should carefully consider selecting variables and assign weights. Besides, this bias could be correct by using one of the above passages' approaches to avoidance or mitigation.

1.3 Random forest

Random forest is another form of using multiple decision-tree models to make a much more complicated decision-making process. Unfortunately, no evidence shows a relationship

between the forest's performance and the number of trees within the forest. It tells that the number of trees may affect the results of the decision tree analysis. It can go either worse or better, and there are no clear signs (Oshiro, 2012). This also means that if the decision-tree model is sufficient for the purpose, keep it as simple as possible, and there's no need to pursue an advanced model (Quinlan, 1999).

2. Conclusions

2.1 Synopsis

Most of the research reviewed concluded that the decision model must be efficient, simple, and functional. The papers used in this review represent lots of essentials about how to create an excellent decision-tree model. However, in contrast to the models they presented in their reports, it is way too complicated for the project required. Still, these reports provide some great insights into an excellent decision-tree model.

2.2 Inform the project

This project needs to build a decision tree model for teaching and providing researchers with a simple analysis. It is not similar to the research model in the literature. Therefore, compared with the literature model, it is relatively simple and straightforward, and only needs modeling can be done according to the traditional decision tree model. On the other hand, the research of random forest belongs to advanced modeling. According to the definition in the section synopsis, it is not mandatory that the project can achieve this goal.

The existing research around the algorithm developed for the decision-tree model is fundamental. The algorithm's depth determines the upper limit of the decision tree model, which can maximize the module's application (Povkhan, 2020). Only select the relative reference

module to ensure the development of the module. While developing the modules, by comparing the modules made by MTU with the literature modules, the optimal structure can be quickly found to clarify the ideas for developing the project modules. Then build new modules based on MTU's module architecture to achieve project expectations.

The literature suggests effective integration of data management processes such as data mining and data classification. With the development of technology today, it is possible to achieve modular management for data management that once required various professionals to complete together (Rupnik, 2007). This means that non-professionals can also process data well through software and tools. It has also been proposed in the literature that data sorting is also a way to improve decision tree results accuracy (Yang, 2013). Since this project's module development is not pursuing extensive data processing capabilities, it is simply applied to teaching, reducing or even not favor algorithms' development. By using publicly existing modules and developing and upgrading, the project can be achieved the requirements and standards.

2.3 Project implementation

The MTU website introduces lots of valuable information about how to make a decision tree model in the Jamovi environment. It stated the ability to pruning and splitting the trees (Chetan, 2018). In addition to the requirements proposed by the project sponsor, instead of making the module too complicated, the module should have options for how many splits are used for the coming analysis. In contrast to other platforms providing trees splitting by clicks, and the new module should take it as an option.

References

Chetan B., Kishen S., Mitul S., Rajasekar K (2018). MTU Classification-Decision tree and random forest. https://sites.google.com/mtu.edu/mtujamovi/mtu-classification-decision-tree-and-random-forest?authuser=0

Jena, M., & Dehuri, S. (2020). Decision tree for classification and regression: A state-of-the-art review. Informatica (03505596), 44(4), 405–420. Doi: https://doi.org/10.31449/inf.v44i4.3023

Kamiński, B., Jakubczyk, M., & Szufel, P. (2018). A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research, 26(1), 135–159. Doi: https://doi.org/10.1007/s10100-017-0479-6

Kam Jacobi, S. K., & Hobbs, B. F. (2007). Quantifying and mitigating the splitting bias and other value tree-induced weighting biases. Decision Analysis, 4(4), 194-210. Doi: https://doi.org/10.1287/deca.1070.0100

Liu, Y., & Salvendy, G. (2007) Visualization support to better comprehend and improve decision tree classification modelling process: a survey and appraisal, Theoretical Issues in Ergonomics Science, 8:1, 63-92. Doi: https://doi.org/10.1080/14639220500284371

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012, July). How many trees in a random forest? In International workshop on machine learning and data mining in pattern recognition (pp. 154-168). Springer, Berlin, Heidelberg. Doi: https://doi.org/10.1007/978-3-642-31537-4_13

Povkhan, I., & Lupei, M. (2020). The algorithmic classification trees. 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP), Data Stream Mining & Processing (DSMP), 2020 IEEE Third International Conference On, 37–43. Doi: https://doi.org/10.1109/DSMP47368.2020.9204198

Quinlan, J. R. (1999). Simplifying decision trees. International Journal of Human-Computer Studies, 51(2), 497-510. Doi: https://doi.org/10.1016/S0020-7373(87)80053-6

Rupnik, R., Kukar, M., & Krisper, M. (2007). Integrating data mining and decision support through data mining based decision support system. The Journal of Computer Information Systems, 47(3), 89-104. Doi: https://doi.org/10.1109/ITI.2006.1708482

Shaheen, M., Zafar, T., & Ali Khan, S. (2020). Decision tree classification: Ranking journals using IGIDI. Journal of Information Science, 46(3), 325–339. Doi: https://doi.org/10.1177/0165551519837176

Yang, H., & Fong, S. (2013). Improving adaptability of decision tree for mining big data. New Mathematics and Natural Computation, 9(1). Doi: https://doi.org/10.1142/S1793005713500063

Appendix A

Project Acceptance Document

Sponsor's Project Acceptance Document

This document is the means by which your project sponsor formally agrees that your project has been satisfactorily completed and that it meets the project goal and objectives that were set at the onset of the project. It is therefore important that you describe the goal, objectives, and related metrics in the appropriate section below. The "PLAN" section is to be prepared at the beginning of the project and the "RESULTS" and "ACCEPTANCE" sections after your project has been completed. Your sponsor should provide input and sign where indicated. The signed document will also be a required section in your Project Final Report. This document is a template whose sections may be expanded as necessary.

Project Name: Decision Tree Analyzer
Student Name: Guangning Wei
Sponsoring Organization: NYU SPS MASY

Project Sponsor Name and Title: Dr. Andres Fortino
Project Sponsor Contact Information (email and phone): cell: 845-242-7614
Email:agf249@nyu.edu

PLAN

PROJECT PLAN

At project start, show the project goal; the project objectives and related metrics to be used to show successful project completion. Sponsor should sign to indicate agreement.

Project Goal_____ The goal of the project is creating a computer-based tool to perform decision tree analysis in the Jamovi R interface program

Objective #1_____ Build a Jamovi add-on to perform decision tree analysis both for numerical and as well as categorical target variables.

Objective #2_____ The creation of a Git Hub repository of the final project files and deliverables.

Objective #3_____ Deliver a ready to install add-on to Jamovi with clear user instructions for installation and use.

Objective #4_____ Test the validity of the algorithm using client-defined data sets.

I agree with the above planned project goal, project objectives, and related metrics.

Andres Fortino _____ 2/25/21
Project Sponsor Signature **Date:**

RESULTS

PROJECT RESULTS

Planned Start Date: February 1st 2021 **Planned End Date:** May 10th 2021
Actual Start Date: February 1st 2021 **Actual End Date:** ___

If actuals differ from planned dates, the revised dates (Actual) are accepted by the sponsor if initialed here: **Sponsor Initials** _____

Project Goal

Was the project goal achieved as planned? **X**Yes ☐No, Reason missed: AGF
If NO, please explain why this is an acceptable deviation. _____ **Sponsor Initials** _____

Project Objective #1: <as shown above in Plan section>
Did the student's project meet this objective with associated measures and metrics as established at project inception? **Objective#1** ☒has or ☐has not been met. **Sponsor Initials** AGF
If not met please explain why this is or is not an acceptable deviation.

Project Objective #2: <as shown above in Plan section>
Did the student's project meet this objective with associated measures and metrics as established at project inception? **Objective#2** ☒has or ☐has not been met. **Sponsor Initials** AGF
If not met please explain why this is or is not an acceptable deviation.

Project Objective #3: <as shown above in Plan section>
Did the student's project meet this objective with associated measures and metrics as established at project inception? **Objective#3** ☒has or ☐has not been met. **Sponsor Initials** AGF
If not met please explain why this is or is not an acceptable deviation.

Project Objective #4: <as shown above in Plan section>
Did the student's project meet this objective with associated measures and metrics as established at project inception? **Objective#4** ☒has or ☐has not been met. **Sponsor Initials** AGF
If not met please explain why this is or is not an acceptable deviation.

Sponsor's Overall Evaluation of student's performance: A <expand, as necessary>

ACCEPTANCE

PROJECT ACCEPTANCE

☐ X Project was completed satisfactorily and is hereby accepted

☐ Project was completed satisfactorily but did not meet all objectives, as shown above. The Project is, nevertheless, accepted.

Andres Fortino _____ 4/27/21
Project Sponsor Signature **Date:**

Guoqiang Wei _____ 04/27/21
Student Signature **Date:**

Appendix B

Project Sponsor Agreement

New York University
MS in Management and Systems
Applied Project
Project Sponsor Agreement

1. Goals of the Program

For Participating Organizations

- Begin relationship with New York University
- Receive help from highly trained NYU graduate student
- Provide internship opportunity for NYU graduate student
- Receive assistance at no cost

For NYU Graduate Students

- Manage and implement a meaningful project aligned with their professional and educational goals
- Hands-on experience interacting with a start-up or operational small business or organization
- Earn credit toward completion of graduate degree by conducting an unpaid Applied Project under the mentorship of an NYU-SCPS professor.

2. Project Sponsor and Student Responsibilities

- Student prepares project planning documents
- Sponsor reviews and approves student's project plan
- Student submits project plan to faculty supervisors for approval
- Student conducts project according to plan
- At predetermined milestones sponsor reviews and approves status reports submitted by student
- Status reports reviewed and evaluated by faculty supervisors to assure student effort and project meet course requirements
- Project sponsor and student participate in periodic project reviews with NYU
- At project completion project sponsor completes evaluation forms
- Student prepares final report

3. Project Selection Process

- Project Evaluation Committee reviews proposed projects
- Projects are:
  - Relevant to MS degree course content
  - Significant to the participating organization
  - Substantial in terms of duration and scope
  - Challenging to the student
  - Capable of being measured against predetermined goals

4. The MS in Management and Systems

Concentrations in:

- Enterprise Risk Management

Students Study Courses in:

- Business Management
- Marketing
- Information Technology
- Financial Management
- Financial Products
- Project Management

Typical Participating Student Profile

- Students selected to participate in this program meet stringent criteria
- Have completed all coursework
- High achievers with highest level GPAs and strong academic credentials
- Highly motivated for success

5. Sponsor and Project Information

Type of Organization	☐ For Profit X Not for Profit				
Name of Organization	NYU School of Professional Studies and the Management and Systems program (MASY)				
Address	7 East 12 Street				
City	NY	State	NY	Zip	
Project Sponsor	First Name	Andres	Last Name	Fortino	
Title	Dr.				
Phone	845-242-7614				
Email	Agf249@nyu.edu				
Web Site	www.nyu.edu				
Type of Business	Higher Education				

Student Name	Guangning Wei
Project Title	Decision Tree Analyzer

Description of Project

This project is about building a decision-tree model in R language, and if could add feature for random forest analysis is a plus. There was a similar module built under Jamovi and programmed with R language before by the Michigan Technological University; however, it is not working properly on the client's running platform. This project is to use the knowledge that I learned from the MASY program at NYU SPS to create a brand-new module with same specifications requirements and make sure it can finally properly run for client. The final deliverables of the project must be able to meet the requirements of the client which are: easy use as a teaching tool, and tool for researching purpose. It also required installation guide and user manual for the module provided as well on the final delivery day.

Estimated Hours of Student Participation	256 hours

Anticipated Results

- Build a Jamovi add-on to perform decision tree analysis both for numerical and well as categorical target variables.
- The creation of a Git Hub repository of the final project files and deliverables.
- Deliver a ready to install add-on to Jamovi with clear user instructions for installation and use.
- Test the validity of the algorithm using client-defined data sets.

Knowledge and expertise student will need to be able to complete the project
R programming R studio proficiency Project management

Will the project sponsor be available for periodic meetings with NYU to review progress, address questions and concerns with the professor supervising the program? *This is a requirement for the program*	X Yes ☐ No
Flexible weekly meetings set up with the project manager	

6. Sponsor Agreement

Students are interns, not professional consultants. NYU is not responsible for the outcomes of projects undertaken by students. Work is on a best-efforts basis; no guarantees or warranties are expressed or implied. Organization is responsible for evaluating work presented, determining its value and whether to use it or not. Some projects may require on-going management or even re-work by the Organization after the student completes their Applied Project.

Please note that in order to post an unpaid position, the internship must encompass all 6 components below:

1. The internship, even though it includes actual operation of the facilities of the employer, is similar to training which would be given in an educational environment;
2. The internship experience is for the benefit of the intern;
3. The intern does not displace regular employees, but works under close supervision of existing staff;
4. The employer that provides the training derives no immediate advantage from the activities of the intern; and on occasion its operations may actually be impeded;
5. The intern is not necessarily entitled to a job at the conclusion of the internship; and
6. The employer and the intern understand that the intern is not entitled to wages for the time spent in the internship.

I have read and agree with the information shown in the Terms and Conditions for employers contained on the following web page(s): http://www.nyu.edu/life/resources-and-services/career-development/employers/post-a-job/terms-and-conditions.html

Please complete and sign this form in the space provided below and return to the course professor via the student who will upload the document to the course drop-box. For any questions, please email the professor: Prof. Israel Moskowitz im36@nyu.edu.

I agree to the all of the above

Participating Organization NYU School of Professional Studies Date 2/25/21

By (signature): Andres Fortino
Project Sponsor

Printed Name: Dr. Andres Fortino

Title: Clinical Associate Professor

7. Student Agreement

Students who are planning to conduct an unpaid Applied Project must read and agree to the "Important Considerations Before Accepting a Job or Internship" contained on the following web page(s): http://www.nyu.edu/life/resources-and-services/career-development/find-a-job-or-internship/important-considerations-before-accepting-a-job-or-internship.html.

Students do not register their Applied Project with the Wasserman Center.

I agree to the all of the above

Student Name (Print) Guangning Wei Date March 2nd, 2021

Signature: Guangning Wei

Appendix C

Project Charter

Decision Tree Analyzer Project Charter

Project Manager: Guangning Wei
Sponsor: Dr. Andres Fortino
Prepared by: Guangning Wei

Name and Location of Client Organization:

Revision History

Revision date	Revised by	Approved by	Description of change

Project Goal

The project aims to create a computer-based tool to perform decision tree analysis in the Jamovi R interface program.

Problem/Opportunity Definition

Complete the development of the Jamovi add-on and make sure it works for both Windows-based PCs and well ad Apple Macs to do decision-trees without type in any R code.

Proposed Project Description

The current Jamovi decision tree add-on developed at Michigan Technological University does not work correctly. The tool is not complete and does not work for both mac and windows platforms. Instead of updating the add-on developed by MTU, this project will create a new version of the project client's add-on demands.

Project Sponsor

- Name and Title:
  Name: Dr. Andres Fortino
  Title: Clinical Associate Professor of Management and Systems, Academic Community of Practice Leader
- Role within the organization: Dr. Andres Fortino is a teaching professional for NYU SPS
- Role on the project: Dr. Andres Fortino runs as project sponsor for this project; he will provide the required information to complete this project.

Objectives:
Technical Objectives:

- Build a Jamovi add-on module in R and make sure that the module works on both mac and windows environments.
- Upload the module into the Github repository.

Timing objectives

- The project must complete before May 10th, 2021, which is before the final presentation.

Resource objectives:

- Utilized the required resources to finish the project on time.

Budget objectives:

- Not identifiable.

Budget objectives:

	Planned	Actual
Salaries	$30,324	No cost(pro-bono engagement)
Documentation	$0	$0
Construction	$0	$0
Mover	$0	$0
Total	**$30,324**	**$0**

Scope objectives:

- Pass the data set test before the final module delivered to the client

Project Selection & Ranking Criteria

Project benefit category:

❑ Compliance/Regulatory X Efficiency/Cost reduction ❑ Revenue increase

Portfolio fit and interdependencies

People who used this module will not have to pay extra money on decision tree analysis.

Project urgency

No immediate urgency; the module could improve the course's efficiency but not urgent to be complete.

Cost/Benefit Analysis

Tangible Benefits

Each student can save $6 per month on the software purchase for decision-tree analysis for two years program to save $142 per student. NYU has more than 50,000 students who saved $7,100,000 for students.

Intangible Benefits

Students are satisfied saving on extra software purchased for decision-tree analysis. Also, professors don't have to look for paid tools to teach decision-tree.

Cost Categories

Personnel cost budget:
-1 person X 288 hours
-Billing rate $70,000 X 1.5 = $105,000/cost X2 = $210,000
-Billing per hour = $210,000/2000hrs = $105/hr
-Cost to the customer for the 288 hours project = $30,240
-Cost per hour of a yearly salary = $80,000/2,000hrs = $40/hr

Other Business Benefits

No other identifiable business benefits.

Scope

- **Quality**
  - The module must work appropriately in Jamovi at both mac and windows.
  - User manual and installation guide must provide.
- **Time**
  - Deliver the module on May 10th, 2021.
  - I have the module upload to the Github repository by May 10th, 2021.

Out of scope activities

- Add random forest analysis

Constraints

1. Creating a program with such a platform required lots of expert knowledge and the ability to adapt the changes such as the packages updates, developing environment changes, and any other external affections. It is hard to continually build the best performance apps right on time, but it will require frequent updates sooner after the project is finished.

2. Decision tree analysis only, can't do any other analysis.

Risks and Mitigation Strategies

Risk: Jamovi updates change the developing environment

Strategies: 1. Make sure the client is aware of the changes

2. Deliver the module work with the old version of Jamovi

Communications Plan

1. Frequency: weekly or biweekly meeting

2. Method: zoom video meeting
3. Content: weekly progress reports, change of the project management, next week's plans.

Schedule Overview

Project Start Date: January 10th, 2021

Estimated Project Completion Date: May 10th, 2021

Major Milestones:

1. Convert the client's discussions into well-defined specification requirements, including all the features and functions for this module.

Deliverable: Delivery the typed-document in mid-February and go back forth with the client to discuss any changes.

2. Build a Jamovi add-on to perform decision tree analysis both for numerical and well as categorical variables.

Deliverable: Delivery as a module by May 10th, 2021.

3. The creation of a Git Hub repository of the final project files and deliverables.

Deliverable: Create the link for Git Hub with all the module files and deliverables to make it public available for anyone to use for decision-tree analysis.

4. Deliver a ready-to-install add-on to Jamovi with clear user instructions for installation and use.

Deliverable: Delivery the installation guide and user manual by May 10th, 2021

5. Test the validity of the algorithm using client-defined data sets.

Measurement: The final product will be delivered by the end of the semester with permission to use by showing the functionalities in the presentation.

Resources Required

Role	Responsibilities	Duration of work	Qualifications needed
Project manager	Leader of the project	12 hours	Project management
Developer	Develop the project's module	276 hours	R programming language

Facilities , Software, Hardware and other Resources

Jamovi: the target software where the module is going to be install

R studio: the developing tool for the module developing process

Procedures/ Methodology

Level	WBS Code	Element Name	Definition	Due By
1	1	Widget Management System	All work to implement a new widget management system.	
2	1.1	Initiation	The work to initiate the project.	
3	1.1.1	Project Determination	Decide which project to take	January 10th 2021
3	1.1.2	Develop Project Proposal	Project Manager to develop the proposal with deliverables.	January 18th 2021
3	1.1.3	Submit Project Proposal	Delivery the proposal to the project sponsor for approve.	February 1st 2021
3	1.1.4	Project Sponsor Agreement and Sponsor Acceptance	Project Manager develops the Project Sponsor Agreement and Sponsor Acceptance	February 25th 2021
3	1.1.5	Project Sponsor Agreement and Sponsor Acceptance Signed/Approved	The Project Sponsor signs Project Sponsor Agreement and Sponsor Acceptance which authorizes the Project Manager to move to the Planning Process.	February 25th 2021
2	1.2	Planning	The work for the planning process for the project.	
3	1.2.1	Project Kickoff Meeting Time Setup	Project Manager set up a time to meet with Project Sponsor and decide weekly meeting time.	February 1st 2021
3	1.2.2	Determine the requirements of the project	The Project Manager determines the project required resource	February 1st 2021
3	1.2.3	Project Kickoff Meeting	The planning process is officially started with a project kickoff meeting which includes the Project Manager and Project Sponsor	February 4th 2021
3	1.2.4	Develop Project Plan	Project Manager the team develops the project plan.	February 5th 2021
3	1.2.5	Submit Project Plan	Project Manager submits the project plan for approval.	February 6th 2021
3	1.2.6	Milestone: Project Plan Approval by Meeting with Project Sponsor	The project plan is discuses between and the Project Manager and Project Sponsor, then the Project manager has permission to proceed to execute the project according to the project plan.	February 11th 2021
2	1.3	Execution	Work involved to execute the project.	
3	1.3.1	Project Meeting with Specification Requirements	The planning process is officially started with a consent on the specification requirements in the meeting between Project Sponsor and the Project Manager	February 11th 2021
3	1.3.2	Install developing software and packages	The project is start with the required software and packages.	February 12th 2021
3	1.3.3	Develop System selection (Mac OS, Linux, Windows)	The project manager decides the developing system for the project between MacOS, Linux Windows.	February 12th 2021
3	1.3.4	Develop System Alternatives If Possible	The project manager decides the alternative developing system for the project between rest of the two after the module works on the previous system working properly.	April 30th 2021
3	1.3.5	Draft Module Finished	The draft module finished.	April 20th 2021
3	1.3.6	Testing Phase	The module is tested with sample data.	May 1st 2021
3	1.3.7	Install the Module on Both Systems	The final module is installed on application for both systems.	May 2nd 2021
3	1.3.8	Video Made for Introduction of the Module	Create a video for demonstration of how to use this module	May 5th 2021
3	1.3.9	Upload to the public as an open source.	Modules upload to the public website and prepare for downloading	May 3rd 2021
2	1.4	Control	The work involved for the control process of the project.	
3	1.4.1	Project Management	Overall project management for the project.	
3	1.4.2	Project Status Meetings	Weekly meetings with reports on the progress of the project.	Every Thursday at 11 am for the meeting if needed.
3	1.4.3	Update Project Management Plan	Project Manager updates the Project Management Plan as the project progresses.	Every Wednesday based on the progress or changes for the project.
2	1.5	Closeout	The work to close-out the project.	
3	1.5.1	Submit Sponsor Acceptance to Sign the Last Part	Project Sponsor signed the last part to agree the accomplishment of the project.	May 10th 2021
3	1.5.2	Update Files/Weekly Summary	All files should be regulated and integrated all the weekly summary.	May 8th 2021
3	1.5.3	Presentation of the Project and Gain formal Acceptance	Giving the presentation of the project and illustrate the module on both systems.	May 10th 2021
3	1.5.4	Upload to GitHub Repository	The files of the module must upload to GitHub repository as open source	May 9th 2021
3	1.5.5	Develop the Installation guide and User manual	Having installation guide and user manual ready to turn in with the module.	May 9th 2021

Project Evaluation

1. Project finish on time
2. Module work on both mac and windows
3. Module has a user manual and installation guide
4. Module passed the data set test
5. Module upload on Github repository and available for download

Appendix D

Project Plan

NYU SCHOOL OF PROFESSIONAL STUDIES

MASTER OF SCIENCE IN MANAGEMENT AND SYSTEMS
Applied Project Capstone
MASY GC- 4100

MEMORANDUM

TO: Dr. Andres Fortino
FROM: Guangning Wei
DATE: March1st 2021

RE: **Assignment 3B – Work Breakdown Structure and Schedule**

Project Tasks Outline

This document is designed to construct work breakdown structure ad schedule for the project of developing decision-tree module for Jamovi environment.

OUTLINE VIEW

1. Widget Management System
   - 1.1 Initiation
     - 1.1.1 Project Determination
     - 1.1.2 Develop Project Proposal
     - 1.1.3 Submit Project Proposal
     - 1.1.4 Project Sponsor Agreement and Sponsor Acceptance
     - 1.1.5 Submit Project Sponsor Agreement and Sponsor Acceptance
   - 1.2 Planning
     - 1.2.1 Project Kickoff Meeting Time Setup
     - 1.2.2 Determine the Requirements of the Project
     - 1.2.3 Project Kickoff Meeting
     - 1.2.4 Develop Project Plan
     - 1.2.5 Submit Project Plan
     - 1.2.6 Milestone: Project Plan Approval by Meeting with Project Sponsor
   - 1.3 Execution
     - 1.3.1 Specification Requirements
     - 1.3.2 Install developing software and packages
     - 1.3.3 Develop System selection (Mac OS, Linux, Windows)
     - 1.3.4 Develop System Alternatives If Possible
     - 1.3.5 Draft Module Finished

1.3.6 Testing phase
1.3.7 Install the Module on Both Systems
1.3.8 Video Made for Introduction of the Module
1.3.9 Upload to the public as an open source.
1.4 Control
1.4.1 Project Management
1.4.2 Project Status Meetings
1.4.3 Update Project Management Plan
1.5 Closeout
1.5.1 Submit Sponsor Acceptance to Sign the Last Part
1.5.2 Update Files/Weekly Summary
1.5.3 Upload to GitHub Repository
1.5.4 Presentation of the Project and Gain formal Acceptance
1.5.5 Develop the Installation Guide and User manual

Work Breakdown Task Definition and Schedule

Level	WBS Code	Element Name	Definition	Due By
1	1	Widget Management System	All work to implement a new widget management system.	
2	1.1	Initiation	The work to initiate the project.	
3	1.1.1	Project Determination	Decide which project to take	January 10th 2021
3	1.1.2	Develop Project Proposal	Project Manager to develop the proposal with deliverables.	January 18th 2021
3	1.1.3	Submit Project Proposal	Delivery the proposal to the project sponsor for approve.	February 1st 2021
3	1.1.4	Project Sponsor Agreement and Sponsor Acceptance	Project Manager develops the Project Sponsor Agreement and Sponsor Acceptance	February 25th 2021
3	1.1.5	Project Sponsor Agreement and Sponsor Acceptance Signed/Approved	The Project Sponsor signs Project Sponsor Agreement and Sponsor Acceptance which authorizes the Project Manager to move to the Planning Process.	February 25th 2021
2	1.2	Planning	The work for the planning process for the project.	
3	1.2.1	Project Kickoff Meeting Time Setup	Project Manager set up a time to meet with Project Sponsor and decide weekly meeting time.	February 1st 2021
3	1.2.2	Determine the requirements of the project	The Project Manager determines the project required resource	February 1st 2021

3	1.2.3	Project Kickoff Meeting	The planning process is officially started with a project kickoff meeting which includes the Project Manager and Project Sponsor	February 4th 2021
3	1.2.4	Develop Project Plan	Project Manager the team develops the project plan.	February 5th 2021
3	1.2.5	Submit Project Plan	Project Manager submits the project plan for approval.	February 6th 2021
3	1.2.6	Milestone: Project Plan Approval by Meeting with Project Sponsor	The project plan is discuses between and the Project Manager and Project Sponsor, then the Project manager has permission to proceed to execute the project according to the project plan.	February 11th 2021
2	1.3	Execution	Work involved to execute the project.	
3	1.3.1	Project Meeting with Specification Requirements	The planning process is officially started with a consent on the specification requirements in the meeting between Project Sponsor and the Project Manager	February 11th 2021
3	1.3.2	Install developing software and packages	The project is start with the required software and packages.	February 12th 2021
3	1.3.3	Develop System selection (Mac OS, Linux, Windows)	The project manager decides the developing system for the project between MacOS, Linux Windows.	February 12th 2021

3	1.3.4	Develop System Alternatives If Possible	The project manager decides the alternative developing system for the project between rest of the two after the module works on the previous system working properly.	April 30th 2021
3	1.3.5	Draft Module Finished	The draft module finished.	April 20th 2021
3	1.3.6	Testing Phase	The module is tested with sample data.	May 1st 2021
3	1.3.7	Install the Module on Both Systems	The final module is installed on application for both systems.	May 2nd 2021
3	1.3.8	Video Made for Introduction of the Module	Create a video for demonstration of how to use this module	May 5th 2021
3	1.3.9	Upload to the public as an open source.	Modules upload to the public website and prepare for downloading	May 3rd 2021
2	1.4	Control	The work involved for the control process of the project.	
3	1.4.1	Project Management	Overall project management for the project.	
3	1.4.2	Project Status Meetings	Weekly meetings with reports on the progress of the project.	Every Thursday at 11 am for the meeting if needed.
3	1.4.3	Update Project Management Plan	Project Manager updates the Project Management Plan as the project progresses.	Every Wednesday based on the progress or changes for the project.

2	1.5	Closeout	The work to close-out the project.	
3	1.5.1	Submit Sponsor Acceptance to Sign the Last Part	Project Sponsor signed the last part to agree the accomplishment of the project.	May 10th 2021
3	1.5.2	Update Files/Weekly Summary	All files should be regulated and integrated all the weekly summary.	May 8th 2021
3	1.5.3	Presentation of the Project and Gain formal Acceptance	Giving the presentation of the project and illustrate the module on both systems.	May 10th 2021
3	1.5.4	Upload to GitHub Repository	The files of the module must upload to GitHub repository as open source	May 9th 2021
3	1.5.5	Develop the Installation guide and User manual	Having installation guide and user manual ready to turn in with the module.	May 9th 2021

Appendix E

Situational Analysis

NYU SCHOOL OF PROFESSIONAL STUDIES

MASTER OF SCIENCE IN MANAGEMENT AND SYSTEMS
Applied Project Capstone
MASY GC- 4100

MEMORANDUM

TO: Dr. Andres Fortino
FROM: Guangning Wei
DATE: March 1st, 2021

RE: **Assignment 3A – Situational Analysis**

1. Industry analysis

Graduate-level education has been on the hit in recent years, and there are many reasons why the students are demanding graduate school education so hard. The students are pursuing the development of their carrier, looking for academic excellence in the fields, pursuing scientific research opportunities by continuous learning, the desire for lifelong learning goals, changing the field of study, and dissatisfaction about what they learned from the undergraduate level (Yazar, 2020). With many of these purposes, college graduates are looking for a graduate school that teaches a program to expert knowledge for themselves. Indeed, in college, the courses are more likely to teach students the basics and prepare students for usual works after graduation. The graduate school can offer concentrations to prepare students to become experts in specific fields they wish to join. Also, graduate schools can offer more practical courses based on demands (Yazar, 2020). Therefore, the competition for graduate schools is intense, the comprehensiveness of the sound program becomes the primary consideration for students to

select. To attract more graduate students and apply to the programs, the graduate school offers various courses to meet most students' desired. Under the awareness of having quality, courses become the main concern for most graduate schools (Yazar, 2020). There's required to have proper tools to assist professors teaching; however, using public software may not meet the profit's perspective. This is because the available software is complicated and may not having specific purposes for course projections; also, some advanced software is expensive, which will increase the economic costs for both students and the schools. Therefore, this project was born by following the trends of having quality teaching courses with specific tools at the lowest costs.

2. Competitors

There are lots of competitive software out there that draws competitions. Such as software like SmartDraw, typing codes to analysis in python, R, excel VBA to demonstrate decision-tree analysis or the similar module in jamovi. There's only a shortlist for some types of way to do such analysis, and each of them has advantages and disadvantages. Starting with the paid software, they are professional use to do decision-tree analysis, and they usually have a company to develop the software by keep adding features. The good thing about paid software like SmartDraw is that they are reliable, and users never worried about its consistency. Also, that paid software can use for many purposes; it covers the most user interfaces those users may use for analysis. The disadvantage of paid software is that price not equal for utilities. Users paid monthly or quarterly to use only the software's precise analysis, and other features were probably never used by users. Primarily used for education purposes, it's hard for students to continue using this software after the course ended. The second type of performing an analysis is typing codes in python, R, and other languages. This does not require users to pay for the services, but it requires the users to be familiar with specific packages and programming skills. The last

competitor for this project is the similar module in the same platform that already exists, and it's more potent than the current project specification requirements. The MTU module in Jamovi can do decision-tree analysis, random forest analysis, and it's free for both platform and module. Unfortunately, the module is no longer updated and can't serve as a proper tool for decision-tree analysis purposes. Once it updates by the development team, it may become a more significant threat for the outcome module's competition by this project. In favor of the free platform and open resource development by the Jamovi environment, it is easy to build a free analysis tool to make it more robust as a project manager. In contrast to the paid software, Jamovi is free for download. It contains various free analyses on it; if the users want other analysis, they may just need to search and download specific modules. Contrast to perform an analysis in R and python, Jamovi modules provide straightforward guidance for users to perform analysis. All they have to do is import data and use the module to solve problems. Thirdly, Jamovi module developers can develop the module by following the website's tutorial because of the open source. This project uses this three convivence to be implemented, easy to use by the users, and easy to develop.

3. Stakeholders

3.1 Satisfied

The project should satisfy the project client's expectations as mentioned in the previous industry analysis, the graduate school that offers this module development to make the course much exciting and practical than other graduate schools. In this case, it requires that this module should actively use instead of having the project client looking for a substitution.

3.2 Manage

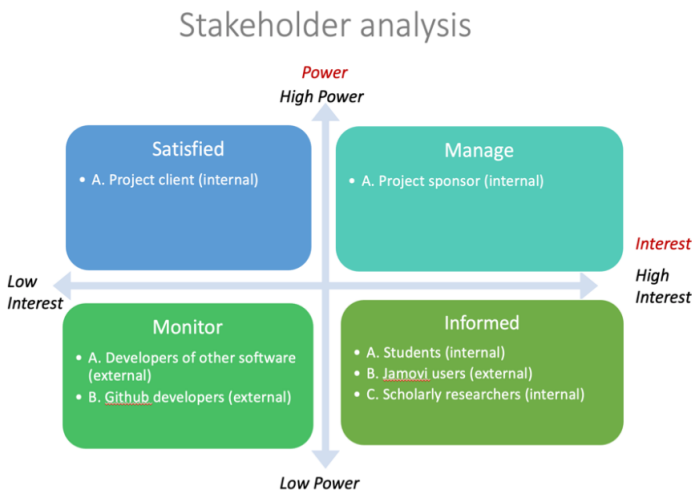
The project sponsor is running connections with the project manager; it is essential to keep them updated on the project's progress and negotiate the possible changes add to the project.

3.3 Monitor

Monitoring is a good approach for having the project's module more competitive in the market. Adding the efficient features from other's modules and software and make it fit with the client's expectations is essential. Also, many GitHub developers had experience developing such tools, and they may offer some ideas for improvements of the project.

3.4 Informed

The final module finished in this project would be used mainly by the students who are in the NYU SPS program, as the project manager, it should be responsible for listening carefully about their thoughts on the final product if possible. Also, even though this module is developed under specific requirements by the project client, but it will finally go to the public platform Jamovi and make it available for the public. So, it is crucial to make the module fit the Jamovi users. At the last stakehold is the scholar research, since the project developed for a graduate school, the module should be in some way to assist the scholarly researchers in their research studies. This is a minor group of stakeholders but is also important for the consideration when develop the module.



4. Porter's five forces

4.1 Threat of potential new entrants

There are many threats for the project's client because of the competition among the graduate schools to provide quality courses to attract students. Software developments for specific purposes mentioned in the industry analysis are not complicated for most graduate schools. For example, graduate schools are financially strong. They may pay other software developments to get what they need specifically. They may just hire software engineering to develop their software with privacy concerns. They may also generate the requirements as projects and offer students academic credits to exchange developments with no costs. The entry-level barrier for this industry is shallow, and each school can find the best way to suit their conditions.

4.2 Bargaining power of buyers and suppliers

The industry customers are students; usually, most graduate schools are offering the course descriptions for their program. Students can review the program and ask the admission office questions to estimate the program's efficiency. Also, the students are likely to choose a well-known graduate school if the program is similar between their possible selections of schools (Das, 2015). The students believe the better school rankings or program ranking can help them develop expert knowledge and future careers. This flows into a double direction of selections. The better graduate school can receive more applications and select the best of them for the graduate program. Students can make their resumes more competitive for applying for jobs by just graduate from this school. Therefore, to make the job acceptance rates better, graduate schools require more knowledgeable professors and practical courses. Another issue coming up with graduate schools heavily invests their money to boost professors' team and strong demand on professional tools to assist teaching. To balance the cost and benefits, the better the school's program ranking will require more tuition to pay by the students. This is the trend of education buyers and suppliers between students and graduate schools.

4.3 Threat of Substitutes

At graduate-level courses, the software uses are distinguishable by the professors' preference. They may ask students to purchase some software or to use some free software to help their teaching. This project on building a module for certain software platforms is an example of having the threat of substitutes. Sometimes the professors believed the paid software is just a tool to help students understand specific knowledge points and may become extra costs. This is because they may only use one semester and end by the semester, less likely for students to continue using these tools after the course ended. Therefore, there's a threat after finishing this

project, if this module is well developed, the professors would likely to accept it for short-term use; otherwise, this project is a failure, and professors may just ask students to use a paid software, even though the students may not use later after the course.

4.4 Degree of competitive Rivalry

As stated in the above sections, there are other tools out there for this project, which is decision-tree analysis. At the software of jamovi, there was one module already built by another team named MTU Classification-Decision tree and random forest (MTU). It is a mature module and functional; however, the lucky thing is that the MTU developers no longer update the module and are not working for the project clients (Chetan, 2018). Currently, the client of the project asks the project manager to develop another module to use. Indeed, the MTU is not the only module that can offer decision-tree analysis; other software already exists in the public that can do the same analysis as well. The project outcome can meet the requirements asked by the client is the main concern of the project manager.

5. Cost benefit analysis

5.1 Cost

Personnel cost budget:

-1 person X 288 hours

-Billing rate $70,000 X 1.5 = $105,000/cost X2 = $210,000

-Billing per hour = $210,000/2000hrs = $105/hr

-Cost to customer for the 288 hours project = $30,240

-Cost per hour of a yearly salary = $80,000/2,000hrs = $40/hr

no other costs occurred.

5.2 Benefits

5.2.1 Tangible benefits.

Each student can save $6 per month on the software purchase for decision-tree analysis for two years program, in total of saving $142 per student. NYU has more than 50,000 students which saved $7,100,000 for students.

5.2.2 Intangible benefits.

Students are satisfied saving on extra software purchased for decision-tree analysis, also professors don't have looking for paid tools to teach decision-tree.

References

Chetan B., Kishen S., Mitul S., Rajasekar K (2018). MTU Classification-Decision tree and random forest. https://sites.google.com/mtu.edu/mtujamovi/mtu-classification-decision-tree-and-random-forest?authuser=0

Das, D. (2015). The grad school admissions statistics we never had. Retrieved February 28, 2021. https://debarghyadas.com/writes/the-grad-school-statistics-we-never-had/

Yazar, T. (2020). Opinions and Suggestions of Graduate Students about Postgraduate. International Online Journal of Educational Sciences, 12(1). https://doi.org/10.15345/iojes.2020.01.009

Appendix F

Risk Management Plan

NYU SCHOOL OF PROFESSIONAL STUDIES

MASTER OF SCIENCE IN MANAGEMENT AND SYSTEMS
Applied Project Capstone
MASY GC- 4100

MEMORANDUM

TO: Dr. Andres Fortino
FROM: Guangning Wei
DATE: Apr 12, 2021

RE: **Assignment 7 – Risk Management Plan**

Project

Decision Tree Analyzer: Creating a decision tree analysis module in Jamovi analysis software

Risks

Number	Risk	Probability Score (1,2 or 3)	Impact Score (1,2 or 3)
1	Jamovi has different capability on mac and windows	3	2
2	The developer may lack the knowledge of programming in R because he used to work with python	3	1
3	Mac probably is not an excellent system to develop the module	2	2
4	The project manager is lacking the knowledge of creating a repository on GitHub for the module	1	1
5			

Risk Matrix

		RISK (exposure)		
		1.Slight	2. Moderate	3. High
Probability (of occurrence)	1. Very Unlikely	4		
	2. Possible		3	
	3. Expected	2	1	

Contingency Plan

Risk	Description	Probability (1-3)	Exposure (1-3)	Contingency Plan
1	Jamovi has different capability on mac and windows	3	2	Ensure the client is aware that the module may run on different versions of Jamovi on mac and windows.
2	The developer may lack the knowledge of programming in R because he is not familiar with R	3	1	
3	Mac probably is not an excellent system to develop the module	2	2	
4	The project manager is lacking the knowledge of creating a repository on GitHub for the module	1	1	
5				

Appendix G

Change Management Plan

PROJECT CHANGE MANAGEMENT PLAN

Project Name:	**Decision Tree Analyzer**
Prepared by:	**Guangning Wei**
Date (MM/DD/YYYY):	**03/28/2021**

To create links to referenced documents (e.g., Link_To_...): *Insert* → *Hyperlink* on your toolbar.
Modify the text of this document to fit your project.

<Modify this template to fit your project.>

1. Purpose

The purpose of this Change Management Plan *is to:*

- Ensure that all changes to the project are reviewed and approved in advance.
- All changes are coordinated across the entire project.
- All stakeholders are notified of approved changes to the project.

All project Change Requests (CR) must be submitted in written form using the Change Request Form provided.	Link_To_Project Change Request Form
The project team should keep a log of all Change Requests.	Link_To_Project Change Request Log

2. Goals

The goals of this Change Management Plan *are to:*

- Give due consideration to all requests for change.
- Identify define, evaluate, approve, and track changes through to completion.
- Modify Project Plans to reflect the impact of the changes requested.
- Negotiate changes and communicate them to project client.

3. Responsibilities

Those responsible for Change Management	*Their Responsibilities*
• Project Manager	Developing the Change Management Plan

3. Responsibilities

Those responsible for Change Management	*Their Responsibilities*
• Project Manager	Facilitating or executing the change management process. This process may result in changes to the scope, schedule, budget, and/or quality plans. Additional resources may be required.
• Project Manager	Maintaining a log of all CRs
• Project Manager	Conducting reviews of all change management activities with senior management on a periodic basis
• Project Manager and client	Ensuring that the *Change Management Plan* is implemented and accepted by both project manager and client

4. Process

<The Change Management process may be simple or complex. The following text is provided as an example of how requests for change can be handled in your project. Supplement with a graphical flowchart if that will help your stakeholders understand the process. Modify as necessary.>

The Change Management process occurs in six steps:

1. Submit written Change Request (CR)
2. Review CRs and approve or reject for further analysis.
3. If approved, perform analysis and develop a recommendation.
4. Accept or reject the recommendation.
5. If accepted, update project documents and re-plan.
6. Notify client of the change.

In practice the Change Request process is a bit more complex. The following describes the change control process in detail:

1. **The client can request or identify a change. He/she uses a *Change Request Form* to document the change request.**
2. **The completed form is sent to a designated member of the Project Team who enters the CR into the *Project Change Request Log*.** Link To Project Change Request Log
3. **CRs are reviewed daily by the Project Manager or designee and assigned one four possible outcomes:**
   - *Reject:*
     - Notice is sent to the submitter.
     - Submitter may appeal (which sends the matter to the Project Manager)
     - Project Manager reviews the CR at its next meeting.

4. Process

- *Defer to a date:*
  - Project Manager is scheduled to consider the CR on a given date.
  - Notice is sent to the submitter.
  - Submitter may appeal (which sends the matter to the Project Team)
  - Project Manager reviews the CR at their meeting.
- *Accept for analysis immediately (e.g., emergency):*
  - An analyst is assigned and impact analysis begins.
  - Project Manager is notified.
- *Accept for consideration by the project team:*
  - Project Manager reviews the CR at its next meeting.

4. **All new pending CRs are reviewed at the Project Team meeting. Possible outcomes:**
   - *Reject:*
     - Notice is sent to the submitter
     - Submitter may appeal (which sends the matter to the Project Sponsor, and possibly to the Executive Committee)
     - Executive Committee review is final.
   - *Defer to a date:*
     - Project Team is scheduled to consider the CR on a given date
     - Notice is sent to the submitter.
   - *Accept for analysis:*
     - An analyst is assigned and impact analysis begins
     - Notice is sent to the submitter.
5. **Once the analysis is complete, the Project Team reviews the results.[1] Possible outcomes:**
   - *Reject:*
     - Notice is sent to the submitter
     - Submitter may appeal which sends the matter to the Project Sponsor (and possibly to the Executive Committee)
     - Executive Committee review is final.
   - *Accept:*
     - Project Manager accepts the analyst's recommendation
     - Notice is sent to Project Sponsor as follows:
       - Low-impact CR – Information only, no action required
       - Medium-impact CR – Sponsor review requested; no other action required
       - High-impact CR – Sponsor approval required.

[1] Note: Sponsor participates in this review if the analysis was done at Sponsor's request.

4. Process

▪	*Return for further analysis:*	Project Manager has questions or suggestions that are sent back to the analyst for further consideration.
6.	**Accepted CRs are forwarded to the Project Sponsor for review of recommendations. Possible outcomes:**	
▪	*Reject:*	• Notice is sent to the submitter • Client review is final.
▪	*Accept:*	• Notice is sent to the submitter • Project Manager updates relevant project documents • Project Manager re-plans • Project Manager acts on the new plan.
▪	*Return for further analysis:*	• The Sponsor has questions or suggestions that are sent back to the analyst for further consideration • Notice is sent to the submitter • Analyst's recommendations are reviewed by Project Manager (return to *Step 5*).

5. Notes on the Change Control Process

<Modify Sections 3 and 5 to meet the needs of your project.>

1. A Change Request is:

- Included in the project only when both Sponsor and Project Manager agree on a recommended action.

2. The CR may be:

- *Low-impact* – Has no material affect on cost or schedule. Quality is not impaired.
- *Medium-impact* – Moderate impact on cost or schedule, or no impact on cost or schedule but quality is impaired. If impact is negative, Sponsor review and approval is required
- High-impact – Significant impact on cost, schedule or quality. If impact is negative, Executive Committee review and approval is required

3. For this project:

- *Moderate-impact* – Fewer than *X* days change in schedule; less than $*XX* change in budget; one or more major use cases materially degraded
- *High-impact* – More than *X* days change in schedule; more than $*XX* change in budget; one or more major use cases lost.

4. All project changes will require some degree of update to project documents:

- *Low-impact* – Changes likely require update only to requirements and specifications documents
- *Moderate- or high-impact* – depending on the type of change, the following documents (at a minimum) must be reviewed and may require update:

Type of Change:	*Documents to Review (and update as needed):*

5. Notes on the Change Control Process

▪ Scope	▪ Scope Statement and WBS
	▪ Budget
	▪ Project Schedule
	▪ Resource Plan
	▪ Risk Response Plan
	▪ Requirements
	▪ Specifications
▪ Schedule	▪ Project Schedule
	▪ Budget
	▪ Resource Plan
	▪ Risk Response Plan
▪ Budget	▪ Budget
	▪ Project Schedule
	▪ Resource Plan
	▪ Risk Response Plan
▪ Quality	▪ Budget
	▪ Project Schedule
	▪ Resource Plan
	▪ Risk Response Plan
	▪ Quality Plan
	▪ Requirements
	▪ Specifications

5. Project documents:

Whenever changes are made to project documents, the version history is updated in the document and prior versions are maintained in an archive. Edit access to project documents is limited to the Project Manager.

- For this project, all electronic documents are kept in (select one of the following and describe it in the adjacent space provided):

[ ] Version Control System: Github

[ ] Central storage available to the Project Team:
Github

[ ] Other:

- For this project, all paper documents are kept in (select one of the following and describe it in the adjacent space provided):

[ ] Project file maintained by the Project Manager:

5. Notes on the Change Control Process

[ ] Other:

- The following individuals have edit access to project documents:

Role	*Documents*
▪ Project Manager	▪ All current documents
	▪ Project archive
▪ Client	▪ All the final documents
▪	▪
▪	▪
▪	▪

6. Project Change Management Plan / Signatures

Project Name:	Decision Tree Analyzer
Project Manager:	Guangning Wei

I have reviewed the information contained in this Project Change Management Plan *and agree:*

Name	Role	Signature	Date (MM/DD/YYYY)
Guangning Wei	Project Manager	GW	03/29/2021
Dr. Andres Fortino	Client	Andres Fortino	4/28/21

The signatures above indicate an understanding of the purpose and content of this document by those signing it. By signing this document, they agree to this as the formal Project Change Management Plan.

Appendix H

Status Reports

<Decision Tree Analyzer> Status Report –March <March 2021>

To: Dr. Andres Fortino **cc:**

From: Guangning Wei

Date: March 28 2021

YOUR ANTICIPATED COMPLETION DATE: May 10th 2021

COMPLETION SEMESTER: Spring,2021 **(e.g. Summer, 2025)**

Project Status Areas:	Execution Week <x>		
	Green	Yellow	Red
1. Overall Project Status			
2. Project Schedule			
3. Project Deliverables			
4. Issues			
5. Project Risks			
6. Resources & Collaboration			
7. Change Status			

**see Assessment Guidelines on the last page of this doc.

1 – Overall Project Status
Status – Overall
• In the last two months, the module has been built with target functionalities. • Project was on schedule and able to be delivered on time. • The module overall is good for the algorithm of calculations, but the graph still needs improve.

2 – Project Schedule	
Tasks that are not on schedule per workplan	Impact
1. N/A	1. N/A

3 – Project Deliverables
COMPLETED DELIVERABLES: • Convert the client's discussions into well-defined specification requirements, including all the features and functions for this module. ***UPCOMING DELIVERABLES:*** • Build a Jamovi add-on to perform decision tree analysis both for numerical and well as categorical variables. • The creation of a Git Hub repository of the final project files and deliverables. • Deliver a ready-to-install add-on to Jamovi with clear user instructions for installation and use. • Test the validity of the algorithm using client-defined data sets.

4 – Issues
N/A

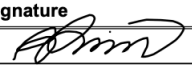
5 – Project Risks	
Potential Risks	**Possible Mitigation**
Jamovi updates change the developing environment, where some packages not able to be downloaded	• Make sure the client is aware of the changes. • Deliver the module work with the old version of Jamovi.

The algorithm is not able to set parameters for the graph.	• Make sure the client is aware of the situations. • Keep working on the current algorithm or change to another algorithm.

6– Resources and Collaboration
• N/A

7 – Change Status	
Scope Changes	**Status** (Requested | Approved | Completed)
N/A	N/A
N/A	N/A

Comments/Actions

8 – Sponsor Signoff		
Name	**Signature**	**Date**
Dr. Andres Fortino		3/29/21

Assessment Guidelines

The assessment is designated by one of the three "Traffic Light" colors utilizing the following guidelines:

Each project should establish the appropriate project slippage metrics for yellow vs red status

Executive Summary:	Assessment		
	Green	Yellow	Red
Overall Project and Most status areas	No major issues, minimal risk to project, on target with expected outcomes, project on schedule, everyone satisfied with progress.	Some major issues, moderate risk to project, must monitor closely, some internal or/and external dissatisfaction with progress. Project plan slipping by 2+ days.	Significant issues, serious risks to project, significant intervention must occur to achieve success, potential for stoppage of project activity. Project slipping by 5+ days, and resources uncommitted to meet deliverables.

In your filename for this document, prefix with Green-, Red-, or Yellow-. G- or R- or Y- and show the date and your name

For example G- Mary_Smith_Nov2099_Status.doc

<Decision Tree Analyzer> Status Report – <April 2021>

To: Dr. Andres Fortino **cc:**

From: Guangning Wei

Date: April 11, 2021

YOUR ANTICIPATED COMPLETION DATE: May 10th 2021

COMPLETION SEMESTER: Spring,2021 (e.g. Summer, 2025)

Project Status Areas:	Execution Week <x>		
	Green	Yellow	Red
1. Overall Project Status			
2. Project Schedule			
3. Project Deliverables			
4. Issues			
5. Project Risks			
6. Resources & Collaboration			
7. Change Status			

**see Assessment Guidelines on the last page of this doc.

1 – Overall Project Status
Status – Overall
• The module has been built with target functionalities. • Project was on schedule and able to be delivered on time. • The module overall is good for the algorithm of calculations, but the graph still needs improve. • The module successfully meets the original deliverables

2 – Project Schedule	
Tasks that are not on schedule per workplan	Impact
1. N/A	1. N/A

3 – Project Deliverables
COMPLETED DELIVERABLES: • Convert the client's discussions into well-defined specification requirements, including all the features and functions for this module. • Build a Jamovi add-on to perform decision tree analysis both for numerical and well as categorical variables. • The creation of a Git Hub repository of the final project files and deliverables. • Deliver a ready-to-install add-on to Jamovi with clear user instructions for installation and use. • Test the validity of the algorithm using client-defined data sets. ***UPCOMING DELIVERABLES:*** • Adding some small features into the module. • Final presentation

4 – Issues
N/A

5 – Project Risks	
Potential Risks	**Possible Mitigation**

6– Resources and Collaboration
• N/A

7 – Change Status	
Scope Changes	**Status** (Requested | Approved | Completed)
N/A	N/A
N/A	N/A

Comments/Actions

8 – Sponsor Signoff		
Name	**Signature**	**Date** 4/11/21
Dr. Andres Fortino		

Assessment Guidelines

The assessment is designated by one of the three "Traffic Light" colors utilizing the following guidelines:

Each project should establish the appropriate project slippage metrics for yellow vs red status

Executive Summary:	Assessment		
	Green	Yellow	Red
Overall Project and Most status areas	No major issues, minimal risk to project, on target with expected outcomes, project on schedule, everyone satisfied with progress.	Some major issues, moderate risk to project, must monitor closely, some internal or/and external dissatisfaction with progress. Project plan slipping by 2+ days.	Significant issues, serious risks to project, significant intervention must occur to achieve success, potential for stoppage of project activity. Project slipping by 5+ days, and resources uncommitted to meet deliverables.

In your filename for this document, prefix with Green-, Red-, or Yellow-. G- or R- or Y- and show the date and your name

For example G- Mary_Smith_Nov2099_Status.doc

Appendix I

Annotated Bibliography

1. Chetan B., Kishen S., Mitul S., Rajasekar K (2018). MTU Classification-Decision tree and random forest.

   Decision tree is a type of supervised learning algorithm (having a pre-defined target variable) that is mostly used in classification problems. It works for both categorical and continuous input and output variables.

   This website including the formal decision tree module built by the MTU group, they had showing the functionalities of the module in jamovi. I think this would be a great resource and guidance for me to build my project.

2. Ho, Tin Kam (1995). Random Decision Forests (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.

   Decision trees are attractive classifiers due to their high execution speed, but tress derived with traditional methods often cannot be grown to arbitrary complexity for possible loss of generalization accuracy on unseen data. The limitation on complexity usually means sub-optimal accuracy on training data. Following the principles of stochastic modeling, we propose a method to construct tree-based classifiers whose capacity can be arbitrarily expanded for increases in accuracy for both training and unseen data. The essence of the method is to build multiple trees in randomly selected subspaces of the feature space. Trees in different subspaces generalize their classification in complementary wats, and their combined classification can be monotonically improved. The validity of the method is demonstrated through experiments on the recognition of handwritten digits.

   This article shows the disadvantages of using a decision tree to process decision tree analysis; then it suggests using the random forest to increase the accuracy of using a single decision tree model. I think this might be useful for building the potential of the future in my paper.

3. Jena, M., & Dehuri, S. (2020). Decision tree for classification and regression: A state-of-the art review. Informatica (03505596), 44(4), 405–420.

   Classification is a way of fitting objects to a category which best suits its characteristics. Classification is a two-step process in which the first one constructs the classifier by examining vividly the training set containing the attributes and their associated class labels [25]. This step is called the training or learning phase [26] [27]. The second step is known as the classification phase where the performance of the classifier is measured for the testing dataset. If performance is found up to the mark, then those rules are applied to unknown data tuples to predict their class labels[28]. Classification intends to distinguish the discrete category of an unknown sample by contemplating a training dataset.

   This journal represents lots of mathematical concepts behind the decision tree running techniques and provides insights into the two main categories of decision tree types: classification and regression. It talks about the algorithm development could empower the

decision tree models and come out with more precise results. I would like to speak about the decision tree model types and algorithm boost techniques in my paper slightly.

4. Kamiński, B., Jakubczyk, M., & Szufel, P. (2018). A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research, 26(1), 135–159.

Sequentially and uncertainty are inherent in managerial practice. The former means that managers have to consider multi-staged strategies, encompassing several actions following one another, rather than only a single step; the latter—that a company's payoffs depend not only on managers' actions but also on exogenous events (states of the world), which may often be perceived as random from the perspective of the decision-maker. The actions and reactions are usually intertwined, further complicating the picture.

This journal represents that the reason for having a decision tree to assist the decision-making process. Usually, the decision-maker would check the probability of success on its actions, but in general, there are the success and fail which is 50% for each; however, due to non-stochastic distributional uncertainty, there is no single expected value that can make by only applying this sense. In my article, I would suggest the decision tree model as an assisting tool to help decision-makers make a better decision after more to concern the random distribution, instead of just measuring each decision either 50% chance to success or 50% chance to fail.

5. Kam Jacobi, S. K., & Hobbs, B. F. (2007). Quantifying and mitigating the splitting bias and other value tree-induced weighting biases. Decision Analysis, 4(4), 194-210.

Multiattribute decision analysis provides a framework for helping decision-makers tackle complex decisions involving conflicting objectives. The decision process often consists in organizing the users' goals into value trees or objectives hierarchies (Keeney and Raiffa 1976, von Winterfeldt and Edwards 1986). Attributes are used in value trees to quantify the extent to which an alternative achieves an objective. The relative importance of the attributes is described by attribute weights, which can be derived either nonhierarchically by simultaneously evaluating all attributes, or hierarchically by assigning weights to subsets of attributes at each level of the tree. Lower-level attribute weights are found by multiplying down the tree.

This article built a model to test the biases when establishing the value tree model, where represents that the different weights on variables would create the form of the tree into several results and describe methods for bias correction. I believe this could definitely assist in my paper for describing decision trees.

6. Liu, Y., & Salvendy, G. (2007) Visualization support to better comprehend and improve decision tree classification modelling process: a survey and appraisal, Theoretical Issues in Ergonomics Science, 8:1, 63-92.

Rapid development in information technologies has changed dramatically the way in which data are collected, stored and accessed. In fact, the growth in the amount of data has far exceeded humans' abilities to capture and analyze without using powerful tools. Therefore, unless advanced knowledge discovery techniques are developed, organizations risk losing much

valuable information from the data they have collected and stored. Data mining (DM) can be defined as the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data (Fayyad et al. 1996).

This journal is trying to tell decision-makers how important it is for the decision tree goes to visualization. It stated that in the big data age, where the data set is holding a large amount of data, without adding visualizations as a feature in the decision tree model, it might not attract decision-makers' attention. This would be a great point in my paper about how to build a good decision tree model and also helps me with the development guide.

7. Mathew, P. V. (1981). Decision tree: An analytical tool for decision making. Decision, 8(1), 29.

A useful framework for analyzing decision problems, however obscured and circumvented it may be in actual practice, could be laid down in terms of: (a)identifying problems, (b)generating alternative courses of action, (c)determining the results of each course of action,(d)evolving criteria to evaluate the alternatives, and selecting the best course of action. Every alternative course of action identified may lead to a series of uncertain events in future. The results of different decisions taken today, in general will not only depend on these uncertain events, but also on the future decisions that will be made after the events occur. Thus, decision maker cannot decide what he should do now unless he takes into account the effects of the uncertain events and the course of action he will choose in future after the events have occurred. Because of these interrelationships among the present alternatives, the uncertain events, and the future decisions, the whole decision-making process becomes very complex.

This article grants basic organizational structure for building a decision tree model, and how to make it with the full potential of doing the works. I think this is an old source from 1981, and it gives some generate insights on how to build a decision tree and what should be expecting from it.

8. Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012, July). How many trees in a random forest. In International workshop on machine learning and data mining in pattern recognition (pp. 154-168). Springer, Berlin, Heidelberg.

Our main conclusions are as the number of trees grows, it does not always mean the performance of the forest is significantly better than previous forests (fewer trees), and doubling the number of trees is worthless. It is also possible to state there is a threshold beyond which there is no significant gain, unless a huge computational environment is available. In addition, it was found an experimental relationship for the AUC gain when doubling the number of trees in any forest. Furthermore, as the number of trees grows, the full set of attributes tend to be used within a Random Forest, which may not be interesting in the biomedical domain. Additionally, data sets' density-based metrics proposed here probably capture some aspects of the VC dimension on decision trees and low-density datasets may require large capacity machines whilst the opposite also seems to be true.

This journal stated that in a random forest, there's having no clear linear relationship between the performance of the forest and the number of trees within the forest. It tells that the number of trees may affect the results of the decision tree analysis. It can go either worse or better, and there's no clear sign. This is a great proven concept that may help my

paper on how to develop a random forest if that is available for the development of the project.

9. Povkhan, I., & Lupei, M. (2020). The algorithmic classification trees. 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP), Data Stream Mining & Processing (DSMP), 2020 IEEE Third International Conference On, 37–43.

   The methods of building recognition systems, based on methods of logical (algorithmic) classification trees (decision trees), do not have this shortcoming. The peculiarity of the logical tree method (the algorithmic classification tree method) is the possibility of complex use for solving each specific task of constructing the recognition scheme of many known algorithms (methods) of recognition. The research is based on the single methodology - the optimal approximation of the training dataset with the help of a set of generalized features (autonomous classification algorithms) that are part of a scheme (an operator) built in the training process.

   According to this paper, it stated that every known algorithms and classification method is restricted to the specialty of the application tasks. This is really related to my paper and project since I am building a decision tree modeling for education purposes which only required the base logic algorithm and made it easier to use for my client, instead of making a giant decision tree for calculating the distance from Earth to Mars. This could also show that how do I skip some of the sophisticated algorithms to build the decision tree model to meet the requirements of clients.

10. Quinlan, J. R. (1999). Simplifying decision trees. International Journal of Human-Computer Studies, 51(2), 497-510.

    Induction algorithms that develop decision trees view the task domain as one of classification. The underlying framework consists of a collection of attributes or properties which are used to describe individual cases, each case belonging to exactly one of a set of classes. Attributes may be either continuous or discrete. A case's value of a continuous attribute is always a real number while its value of a discrete attribute is one of a small set of possible values for that attribute. In real-life tasks it is also important to recognise that a case may have unknown values for one or more of the attributes. A decision tree may be either a leaf identified by a class name, or a structure of the form...

    This journal states the bases structure of making a simple decision tree with mathematical approach, It is really hard to understand, but in my paper, this could prove that the chance for making a simple decision tree is highly possible. It doesn't require the developers or decision makers to be well known about the math and programming. They can still get their hands on making decision tree development by using other tools.

11. Rupnik, R., Kukar, M., & Krisper, M. (2007). Integrating data mining and decision support through data mining based decision support system. The Journal of Computer Information Systems, 47(3), 89-104.

    Data mining can be sued through two approaches. The first approach is called data mining through adhoc data mining projects by the use of data mining software tools… Data mining software tool approach has a disadvantage in a number of various experts needed to collaborate in a project , the complexity of software tools and in transferability of results and models. The disadvantages mentioned call for different approach, which in this paper we call data mining application system approach. Data mining application systems approach signifies the possibility

to develop decision support systems which use data mining methods and do not demand expertise in data mining for business users…

In this journal, where two approaches for data mining were compared. The reason I would like to add this to my paper is because of the trends in the data management process. In the journal, it suggests that the data analysis could be divided into stages: data mining, data collection, input data, output data analysis, along with the development of software and applications. Right now, it doesn't require that many experts in each stage help the data analysis. It is much easier for some non-experts decision-makers to do the decision-making process.

12. Shaheen, M., Zafar, T., & Ali Khan, S. (2020). Decision tree classification: Ranking journals using IGIDI. Journal of Information Science, 46(3), 325–339.

The analysis of ever increasing, huge volume of data that is usually collected from a variety of sources and stored in massive data repositories poses a stern challenge for the decision-making task. To counter this demanding challenge, different seminal techniques have been proposed for the extraction of patterns and knowledge of interest from data of large size for the purpose of decision-making. The extraction of useful knowledge and patterns from large data sets using different techniques is known as data mining. Data mining techniques are popular, and there are numerous applications where these techniques are actively used, like medical imaging [1], driver fatigue detection [2], face recognition [3] and many more.

This journal gives another implementation way called IGIDI, where considers information gain, Gini index, and diversity index into a decision model. This method is for selecting an attribute for decision tree classifier, where added the IGIDI to boost the decision tree accuracy and time sensitivity. Compare to another article in this reference list that adding features about information gain only. These two articles all presented the decision tree's diversity where can build anything the decision-makers want.

13. Yang, H., & Fong, S. (2013). Improving adaptability of decision tree for mining big data. New Mathematics and Natural Computation, 9(1).

When sufficient statistics have accumulated at each leaf, a node-splitting algorithm determines whether there is enough statistical evidence in favor of a node-split, which expands the tree by replacing the leaf with a new decision node. VFDT learns by incrementally updating the tree while scanning the data stream on the fly. This powerful concept is in contrast to a traditional decision tree that requires the reading up of a full dataset for tree induction. The obvious advantage is its real-time data mining capability, which frees it from the need to store up all of the data to retrain the decision tree because the moving data streams are infinite.

This article describes a new form of a decision tree that evolved from the traditional decision tree due to big data development. It represents that the conventional decision tree required to have complete dataset before entering the generating process, but this new algorithm could help decision-makers quickly following the trends when keep adding more data into the data set, which is more convenient and assist in building decisions along with the changes of the data environment.