

ITBW41 Predictive Analytics Project Guide

2022 Semester 1

Table of Contents

1 Introduction	2
1.1 WIU Synopsis	2
1.2 WIU Learning Outcomes	2
2 Project Scope & Scenario	2
2.1 Project Background & Scope	2
2.2 Project Requirements	3
Key Task 1: Determine Business Objectives	3
Key Task 2: Data Collection and Exploration	4
Key Task 3: Data Modification	4
Key Task 4: Modelling	4
Key Task 5: Report	4
2.3 Roles & Responsibilities	4
2.4 Project Schedule & Submission	5
3 WIU Assessment Components	6
Annex A: Project Assessment Rubrics	7
Project Proposal (15%)	7
Milestone Report (15%)	7
Presentation (20%)	8
Final Report (30%)	8
Annex B: Submission Report Formats	11
Project Proposal (15%)	11
Milestone Report (15%)	11
Final Project Report (30%)	11

1 Introduction

1.1 WIU Synopsis

Predictive analytics generates future insights with a significant degree of precision based on historical data. Through this WIU, learners will demonstrate their competencies in developing a predictive analytics project in teams by applying data privacy and ethical principles in the collection, use and disposal of data. With the data collected, learners will demonstrate their competencies to impute data format, transform and reshape the business data before applying relevant predictive modelling techniques to predict the desired business outcomes. Learners will then develop data stories for an effective narrative and visual representation of their predictive analytics project.

1.2 WIU Learning Outcomes

Upon completion of this WIU, learners will be able to:

At the end of this CU, learners will be able to:

1. Collect data from multiple sources using appropriate collection tools and techniques that comply with data and privacy ethics.
2. Perform data pre-processing techniques to impute data, transform, reshape and protect the data in accordance with the business requirements and data protection principles.
3. Apply relevant predictive modelling techniques to predict the desired business outcomes to meet the service expectation of the key stakeholders.
4. Work collaboratively in a team to develop dashboards using data storytelling approach for an effective narrative and visual representation of their predictive analytics project.

2 Project Scope & Scenario

2.1 Project Background & Scope

Predictive analytics plays an important role in helping organizations make critical business decisions based on insights and information derived from collected data. Today, given some conditions, businesses want to know what is mostly likely to happen and identify actions that maximizes desirable outcomes. This is taking place across multiple industries, for example, healthcare, manufacturing, finance, retail, marketing manufacturing, human resources, cyber security and education.

Examples of applications of predictive analytics is plentiful. For example, it allows healthcare providers to identify high-risk patients quickly to provide care in a timely manner. Manufacturers can minimize equipment failures by predicting the time each piece of equipment is likely to fail and then service them before it happens. Financial institutions can predict clients who are likely to pay their bills late and can provide timely reminders. Computer system administrators use it to detect intrusions on networks to keep out hackers. Human resource uses it to increase chances of recruiting suitable candidates or reduce attrition.

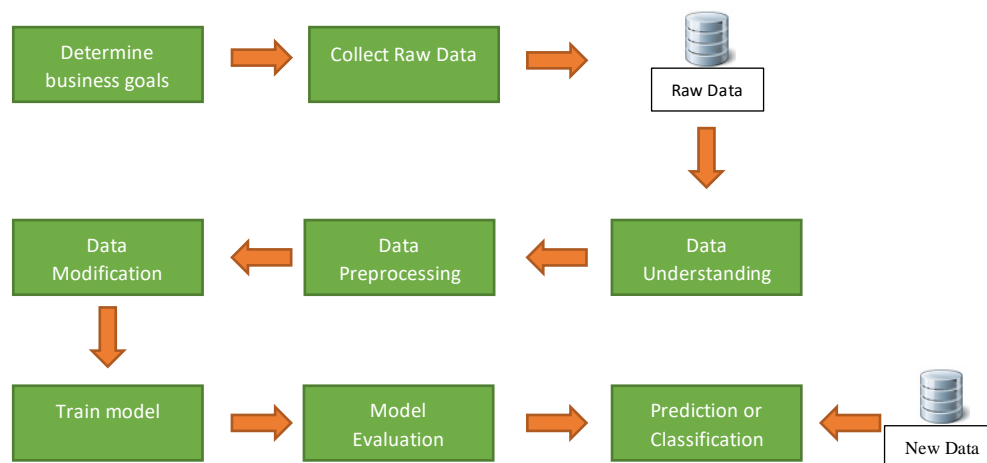
In this project, you are to develop a predictive analytics model using the skills you have acquired in the other units:

- data collection/wrangling to develop effective data storytelling visualization and dashboards.
- Predictive models using supervised and/or unsupervised learning techniques.
- Protection of data privacy.

2.2 Project Requirements

In this project, we will be developing a **classifier**. You will need to identify probable business goals objectives, collect data and use predictive analytics to help meet the business goals.

The suggested process is as follows:



Key Task 1: Determine Business Objectives

As a team, discuss and determine the business goal and objectives. Some examples include:

- To reduce the number of bad loans in a financial institution.
- To keep customers from leaving for competitors' services.
- To improve the efficient in providing health care services.

You can also indicate other objectives like

- To determine the important factors that contribute to an equipment failure.
- To determine the factors that likely to cost our customer to go to the competitors.

You will also need to state the success criteria. This should be measurable. For example:

- We should be able to reduce the number of customers leaving by 10%.
- We should be able to identify 70% of the bad loans.

Plan on how to carry out the project. You should identify the main and sub-tasks and delegate the tasks to each of the members. Also, specify the techniques and tools to be used.

The problem statement should be in a discipline other than IT. For example, it can be in healthcare, education, or finance. You can also consider reaching out to professionals or end users for their inputs on pain points they have experienced in order to formulate your business objectives.

Key Task 2: Data Collection and Exploration

Once the business objectives are defined, you will need to collect data that is relevant to achieving the objectives. Collect data using techniques that comply with data protection and privacy ethics.

Explore the data and understand the characteristics. It would be helpful to use statistical or data visualization techniques to gain a good understanding of the data that you have collected.

You may have to collect more data if you decide that data is insufficient after exploration.

You can also apply unsupervised learning techniques like clustering or association rules to gain better insights into your data.

Key Task 3: Data Modification/Preparation

In this task, you are to prepare the data so that it can be used for modelling. Poor quality data will often result in poor performing models. Examples of steps that you can apply to your data includes imputation, transformation, integration, format, balancing, feature selection and outlier detection/removal, dimensionality reduction.

Key Task 4: Modelling

In the modelling task, the team should apply relevant predictive modelling techniques on the processed data. You can build multiple models using various machine learning algorithms. You should also assess the performance of trained models and tune the hyper-parameters to achieve the best possible outcome. Select the best performing model among the trained models.

Do ensure that you properly partition your data and evaluate your final model. Assess if the model meets the business objectives.

Key Task 5: Report

Write a report, in Microsoft Word format, to document all the tasks that you have done and decisions that you have made in the project. You should document steps that you have taken even if they yield poor or no results. For example, document what you have done to find outliers even if at the end, you are unable to find any. Make sure that your report clearly demonstrates your understanding of the subject matter.

It is important to provide useful interpretation of outcomes and explain all the decisions the team has taken to improve the model.

Include a cover page with your module group number, names and admin number in your report.

Note

It is not the absolute performance (accuracy etc) of your model that matters. Instead, you need to demonstrate that you are doing the correct steps, able to correctly interpret the results generated by the software and make correct decisions or recommendations based on the results.

2.3 Roles & Responsibilities

You are to form a group of 3 to 4 members and elect a group leader.

Project Supervisor / Module Tutor

Provide feedback to project teams' proposals.

- Advise the project group during project development, including the task allocation, development plan, tools used, etc.
- Provide inputs from the user perspective to the project team.
- Act as technical advisors (supervisors are NOT expected to debug programs or solve problems for students).

Group Leader

- Maintain the project schedule and documentation, and record attendance at meetings.
- Coordinate and assign work among members.
- Arrange meetings with the project supervisor. Leader is required to book the meeting rooms if meeting is held outside of the lab/tutorial hours.
- Communicate with the supervisor on behalf of the team to discuss any project-related matters.

Group Members

- Carry out allocated tasks.
- Be punctual and attend all lessons and project meetings.
- Cooperate with the rest of the group during project development.

2.4 Project Schedule & Submission

This project will be completed over 8 weeks, please refer to the following corresponding schedule.

Week	Tasks	Submission/Assessment
Week 11	<ul style="list-style-type: none"> - Understand project requirements - Form project team. - Discussions on project proposal. 	Submit the member via online form (https://forms.gle/YFzwHtWvaJXuJ8uo6)
Week 12	<ul style="list-style-type: none"> - Finalize project proposal - Data Collection - Explore Data. Search for relationship, trends and gain understanding and ideas. - Data Visualization 	- Submit project proposal via Brightspace (15%)
Week 13	<ul style="list-style-type: none"> - Modify Data (Create, select and transformation data in preparation for modelling. e.g. Outlier detection, variable selection, data imputation, sampling and data partitioning.) 	
Week 14		
Week 15	<ul style="list-style-type: none"> - Modelling (Generate models, perform assessment and tuning of the model.) 	(Week 16) - Submit milestone report (15%) and resume (20%)
Week 16		
Week 17		
Week 18	<ul style="list-style-type: none"> - Presentation - Report preparation 	Presentation (20%)
Week 19	Prepare Report/Resume	- Report, Source Codes submission (30%)

Late submissions will be penalised.

*If you do not attend the presentation for week 18, you will not have a chance to explain and clarify your work. This will adversely affect your grades in the report as well losing all the presentation marks.

3 WIU Assessment Components

Assessment Component	Group	Individual	Total
Project Proposal	15%	-	15%
Milestone Report	5%	10%	15%
Final Presentation	10%	10%	20%
Resume	-	20%	20%
Report/Source Codes	10%	20%	30%
Total			100%

Annex A: Project Assessment Rubrics

Project Proposal (15%)

Criteria	Proficient	Competent	Functional	Developing
Business Objectives (4%)	Proposed business objectives/use case/ solution mostly feasible, realistic and reflects the quality and understanding of the subject.	Proposed business objectives/use case/ solution mostly feasible and realistic.	Proposed business objectives/use case/ solution is moderately feasible and realistic.	Proposed business objectives/use case/ solution seems imaginary and impractical.
Project Planning (5%)	Determines resources (i.e. sources of funding, platforms/facilities, consultants/expertise) required for the project development	Outlines the development process with milestones and timeline for the project development	Prepares a project brief/outline to introduce the idea(s), describe the value of the project, and impact to users and stakeholders	Follows an ad-hoc and unstructured project development process
Empathy (2%)	Selects appropriate I&E tools for the empathy study, and examines the pain points experienced by users	Uses suggested I&E tools in a proficient manner for an empathy study, and identifies pain points of users without needing additional guidance	Follows instructions to use suggested I&E tools for an empathy study, and identifies pain points of users through a guided process	Tries one or more I&E tools during a learning activity
Transdisciplinary Thinking (2%)	Collaborates with people from other domains or disciplines to develop idea(s)	Applies an approach used by another domain or discipline to develop idea(s)	Examines approaches used by another domain or discipline to develop a similar idea	Focuses on a singular or own perspective to develop idea(s)
Pitching (2%)	Establishes connections with professionals and/or experts in the field to consult and provide advices on the development and deployment of the product/solution	Obtains inputs and advices from people outside the team who has experiences or insights on the product/solution	Identifies partners or collaborators who the team can approach for help to develop the product/solution	Lists open resources (i.e. coaches, staff with expertise, facilities, small project funding etc) that the team can readily access to develop the product/solution

Milestone Report (15%)

Criteria	Proficient	Competent	Functional	Developing
Data Collection (3%)	Data collected are from relevant sources representing and compressive and not bias.	Data collected are from relevant sources but has limited points of view/approaches.	Data collected are from irrelevant sources and has limited points of view/approaches.	Minimal data collected and are mostly irrelevant.
Data Understanding (6%)	Relevant information from the data is well presented via effective charts, tables, formulae, offering much further and deeper understanding of the data. Provides accurate and thorough description of	Most of the relevant information from the data is presented via charts, tables, formulae, offering good understanding of the data. Provides accurate and good description of the	Some of the relevant information from the data is presented via charts, tables, formulae, offering some understanding of the data. Provides quite accurate and sufficient description	Little of the relevant information from the data is presented. Minimal efforts in the use of charts, tables, formulae, offering no better understand of the data. Provides inaccurate or wrong description of the

	the trends, patterns and insights of the data.	trends, patterns and insights of the data.	of the trends, patterns and insights of the data.	trends, patterns and insights of the data.
Data Preparation (6%)	Performed thorough and comprehensive cleansing and transformation of the data to ensure high quality data ready for modelling.	Performed most of the required cleansing and transformation of the data resulting in good quality data for modelling.	Performed some of the required cleansing and transformation of the data resulting in data with adequate quality for modelling.	Little or no cleansing and transformation of the data resulting in poor quality that are not ready for modelling.

Presentation (20%)

Criteria	Proficient	Competent	Functional	Developing
Visual Aid (5%)	Well designed and effective use of visual aids (e.g. slides) that adds significant values and used seamlessly.	Thoughtful use and helpful in providing clear explanation.	Typical approach, has some value but sometimes hinders clear explanation.	Questionable use and often hinders the explanation and does not value add to the presentation.
Organization (5%)	Well organize and represent information clearly and accurately. Anticipates questions and actively address doubts.	Organized and easy to follow. Provides sufficient information without prompt.	Organized but not easy to follow. Explanation raises questions and doubts.	Disorganized and difficult to follow. Explanation raises questions and doubts and unable to provide convincing answers.
Pace (5%)	Time allocated well, well paced, finishes on time. The explanation was interesting, engaging and provided clear reasoning for the decisions.	A bit too fast or too slow, finishes on time. The explanation provided clear reasoning for the decisions.	A bit too fast or too slow. Overrun. The explanation provided some reasoning for the decisions.	A bit too fast or too slow. Continues with no control of timing. Overrun significantly. The explanation failed to express reasoning for the decisions.
Mannerism (5%)	Engage audience, eye contacts continuously. Confident and presents without referring to scripts.	Appear comfortable, Engage audience. Presents from memorized scripts. Check slides occasionally.	Appear uncomfortable, Some engagement of audience. Frequent checking of slides or script.	Disconnected from audience. Conveys nervousness. Reads from scripts or slides.

Final Report (30%)

Criteria	Proficient	Competent	Functional	Developing
Introduction, Background (3%)	Concisely and clearly describes the project and the process of using analytics to solve the problems	Clearly describes the introduction and background of the subject	Briefly describes the introduction and background of the subject	Confusing description on the subject introduction and background
Modelling (12%)	Uses the many algorithms and perform thorough tuning for the best hyper-parameter values. Good comparison of wide number of well-tune models and conclusively	Uses the some algorithms and perform tuning for the good hyper-parameter values. Able to choose the best performing model.	Uses limited number algorithms, hardly any tuning to get good hyper-parameter values. Able to choose the best performing model.	Uses only one algorithm with hardly any tuning. Model is not necessary the best performing model

	derive the best performing model.			
Coding (3%)	Source codes are working flawlessly, very organized and with detail comments providing explanation for complex steps. Codes are well structured and follows good programming practices.	Source codes are working with some bugs, organized and properly commented. Codes are clear. Follows good programming practices.	Source codes are working with some bugs, disorganized and minimal comments. Codes are unclear and difficult to follow. Follows some good programming practices.	Source codes are buggy, not working, disorganized and without comments. Hardly follows any good programming practices.
Validation (3%)	Uses the appropriate performance metric for the model. Uses proper approaches to ensure validation results are reliable.	Uses some performance metric for the model. Uses proper approaches to ensure validation results are reliable.	Uses some appropriate performance measurement for the model.	Uses inappropriate performance metric for the model.
Process (6%)	Provides a clear rational and convincing reasons for each of the steps in the process from data understanding to final model evaluations.	Provides rational and reasons for most of the steps in the process from data understanding to final model evaluations.	Provides some rational reasons for some of the steps in the process from data understanding to final model evaluations.	No explanation nor reasons provided for each of the steps performed in the process from data understanding to final model evaluations.
Conclusion (3%)	Provides meaningful conclusions related to the business objectives. Useful insights derived from the results are provided.	Provides conclusions related to the business objectives. Some insights derived from the results are provided.	Provides conclusions. Provides some explanation of the results obtained.	No meaningful conclusion, no explanation of the results obtained.

Resume and Cover Letter (20%)

Criteria	Excellent	Very Good	Good	Satisfactory	Needs Improvement
Résumé – Content, Organisation & Language (10%)	<p>Résumé includes all essential sections and contents, all appropriately elaborated on</p> <p>Résumé sections and contents are in appropriate order</p> <p>Résumé shows suitability / strengths (KSAs) well</p> <p>Résumé utilizes STAR formula for one or more projects / items well</p> <p>Résumé language is clear, concise and error free</p>	<p>Résumé includes essential sections and contents, with adequate elaboration</p> <p>Résumé sections and contents are generally in appropriate order, with minor slips</p> <p>Résumé shows suitability / strengths (KSAs) fairly well</p> <p>Résumé utilizes STAR formula for at least one project / item, with minor slips</p> <p>Résumé language is clear and error free</p>	<p>Résumé includes most essential sections and contents, a few of which may lack elaboration / details</p> <p>Most résumé sections and contents are in appropriate order, with some slips</p> <p>Résumé attempts to show suitability / strengths (KSAs), with some slips / misses</p> <p>Résumé utilizes STAR formula for at least one project / item, with some slips</p> <p>Résumé language is clear and largely error free</p>	<p>Résumé includes some essential sections and contents, some of which may lack elaboration / details</p> <p>Some résumé sections and contents are in appropriate order, but needs improvement</p> <p>Résumé attempts to show suitability / strengths (KSAs), with several slips / misses</p> <p>Some attempts at utilizing the STAR formula, although verbs / quantifiers may not be well used</p> <p>Résumé language is comprehensible but</p>	<p>Résumé has missing essential sections and contents; elaboration may be patchy</p> <p>Résumé section and contents are not in appropriate / meaningful order</p> <p>Résumé hardly shows suitability / strengths (KSAs)</p> <p>No attempt at utilizing the STAR formula</p> <p>Résumé contains many errors in word choice, grammar and/or spelling</p>

				contains numerous errors	
Resume – Format (5%)	<p>Résumé is clean, professional and easy to read</p> <p>Style used for headings and body text shows hierarchy of information clearly and is consistent</p>	<p>Résumé is clean and easy to read, but may not look professional</p> <p>Style used for headings and body text is consistent although hierarchy of information could be clearer</p>	<p>Résumé is easy to read for the most part, but some parts may look cluttered</p> <p>Some inconsistencies in the style used for headings and body text</p>	<p>Layout can be improved for better readability and understanding in some places</p> <p>Several inconsistencies in the style used for headings and body text.</p>	<p>Readability can be improved; layout may make some parts difficult to understand</p> <p>Style used for headings and body text lacks consistency</p>
Cover Letter (5%)	<p>Cover letter is well written, highlights suitability and generates interest in the résumé</p> <p>Cover letter utilizes STAR formula for one or more projects / items well</p>	<p>Cover letter shows some suitability and generates some interest in the résumé</p> <p>Cover letter utilizes STAR formula for at least one project / item, with minor slips</p>	<p>Cover letter shows some suitability and mentions the résumé</p> <p>Cover letter attempts to show suitability / strengths (KSAs), with some slips / misses</p>	<p>Cover letter does not show suitability but mentions the résumé</p> <p>Cover letter attempts to show suitability / strengths (KSAs), with several slips / misses</p>	<p>Cover letter does not show suitability and does not mention the résumé</p> <p>No attempt at utilizing the STAR formula</p>

Annex B: Submission Report Formats

Project Proposal (15%)

The proposal should include the followings:

1. Team members (indicate the leader).
2. Introduction/Background information.
3. Business scenario.
4. Business objectives identified.
5. Main and sub tasks identified.
6. Delegation of tasks/responsibility.
7. Project schedule and task allocation (Gantt chart)

Milestone Report (15%)

The report serves as a checkpoint to ensure that you are on track to complete your project. At the time of submission, you should have completed data collection, data understanding and data preparation.

Prepare the report that includes the following content:

1. Team members (indicate the leader).
2. Section on data collection. Indicate the sources of the data as well as the tools and techniques used. Discuss about ethic and data privacy protection.
3. Section on data understanding. Document the steps done to understand the data. You should include any visual aids - charts, graphs, tables, formulae etc. that helped you gain a good understanding of the data.
4. Section on data preparation. This includes all the steps that you have taken to ensure that the data is ready for modelling. You can include, but not limited to, tasks like clustering, outlier detection, data balancing, partitioning, variable selection etc.

It is important that the report demonstrates that you understand the steps necessary to ensure that you have good quality data that is well prepared for training models. You should include aspects of data that you have checked but requires no actions. For example, you should document the steps you have taken to ensure the data is balanced even though at the end you decided that no balancing is required.

Final Project Report (30%)

This report should follow the guidelines stated below. As a rule, your report (excluding the cover page, contents page and the appendixes) **should not exceed 30 pages**. The report should be prepared using **one and a half line spacing**.

Header and/or footer should be included to show at least the project name, the chapter name, the file name of the document and the page number. Each section/chapter should begin on a new page with an appropriate heading. **Name of student** who has documented the section **should appear at the header of the section**.

You may include additional sections such as preface, acknowledgement, glossary of terms and symbols used, etc.

Recommended content outline:

- i) **Cover Page**, this should reflect:
 - WIU Name, Document Title and Supervisor's Name
 - The Project Team Number, Team Member's Name and Admin No. (State roles played, e.g., Team Leader etc.)
 - The Latest Report Amendment Date
- ii) **Contents Page**, this should list all the chapters/sections found in your report with their respective page number.
- iii) **Executive Summary**
 - Summary of business objectives
 - Describe the business scenarios, recommendation and insights discovered
- iv) **Modelling**
 - Describe the modelling techniques used, the tuning process as well as the validation process.
 - Interpret and discuss the results obtained, the implications and actions taken based on the results of modelling to ensure the best possible model is obtained.
 - Describe how you validate the model and ensure that the results you obtained is reliable.
- v) **Conclusion**
 - Discuss if results obtain is able to meet your business objectives. If it does not, provide the reasons and possible course of corrective actions to take.