DEPARTMENT OF ENGINEERING MATHEMATICS

Comparative Machine Learning Analysis for Student Dropout Prediction in a Virtual Learning Environment

Incorporating Student Engagement and Socio-Economic Features

Carlos Duran Calle

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Master of Science in the Faculty of Engineering.

Wednesday 20th August, 2025

Supervisor: Dr. Felipe Campelo

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of MSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Carlos Duran Calle, Wednesday 20^{th} August, 2025

Contents

List of Figures

List of Tables

List of Algorithms

‘

Abstract

A compulsory section, of at most 1 page

This section should summarise the project context, aims and objectives, and main contributions (e.g., deliverables) and achievements. The goal is to ensure that the reader is clear about what the topic is, what you have done within this topic, *and* **what your view of the outcome is.**

Essentially this section is a (very) short version of what is typically covered in more depth in the first chapter. If appropriate, you should include here a clear statement of your research hypothesis. This will obviously differ significantly for each project, but an example might be as follows:

> My research hypothesis is that a suitable genetic algorithm will yield more accurate results (when applied to the standard ACME data set) than the algorithm proposed by Jones and Smith, while also executing in less time.

The latter aspects should (ideally) be presented as a concise, factual list of the main points of achievement. Again the points will differ for each project, but an might be as follows:

- I spent 120 hours collecting material on and learning about the Java garbage-collection sub-system.
- I wrote a total of 5000 lines of *Python* source code, and associated orchestration scripts.
- I designed a new algorithm for computing the non-linear mapping from A-space to B-space using a genetic algorithm.
- I implemented a version of the algorithm proposed by Jones and Smith (2010), corrected a mistake in it, and compared the results with several alternatives.

Supporting Technologies

A compulsory section, of at most 1 page

This section should present a detailed summary, in bullet point form, of any third-party resources (e.g., hardware and software components) used during the project. Use of such resources is always perfectly acceptable: the goal of this section is simply to be clear about how and where they are used, so that a clear assessment of your work can result. The content can focus on the project topic itself (rather, for example, than including “I used LaTeX to prepare my dissertation”); an example is as follows:

- I used the *Pandas* and *Seaborn* public-domian Python Libraries.
- I used a parts of the OpenCV computer vision library to capture images from a camera, and for various standard operations (e.g., threshold, edge detection).
- I used Amazon Web Services for remote storage and processing of data. Specifically, I used:
  - Simple Storage Service (S3) for data storage
  - Elastic Compute Cloud (EC2) for provision of virtual machines
  - Elastic Beanstalk for scaling and load management
  - Sagemaker for all the machine learning components of my project.
- I used LaTeX to format my thesis, via the desktop service *TeXstudio*.

Notation and Acronyms

SDP	:	Student Dropout Predictor
MOOCs	:	Massive Open Online Courses
VLEs	:	Virtual Learning Environments
OULAD	:	Open University Learning Analytics Dataset
ML	:	Machine Learning
RF	:	Random Forest
LG	:	Logistic Regression
LightGBM	:	Light Gradient-Boosting Machine
NNs	:	Neural Networks
SE	:	Student Engagement
KNN	:	K-Nearest Neighbors
KNN	:	K-Nearest Neighbors
KNN	:	K-Nearest Neighbors
KNN	:	K-Nearest Neighbors
KNN	:	K-Nearest Neighbors
KNN	:	K-Nearest Neighbors
x_i	:	the i-th bit of some binary sequence x, st. $x_i \in \{0, 1\}$

Acknowledgements

An optional section, of at most 1 page

It is common practice (although totally optional) to acknowledge any third-party advice, contribution or influence you have found useful during your work. Examples include support from friends or family, the input of your Supervisor and/or Advisor, external organisations or persons who have supplied resources of some kind (e.g., funding, advice or time), and so on.

Dave Cliff writes here to say huge thanks to his colleague Dr Dan Page for sharing this LaTeX thesis template, which was originally written by Dan, for Computer Science dissertations. Dave edited Dan's original to better suit the needs of the Data Science MSc: please don't hassle Dan about any of this, but do feel free to contact Dave if you have any questions or comments on it.

Chapter 1

Introduction

The rise of online learning environments, such as Massive Open Online Courses (MOOCs) and Virtual Learning Environments (VLEs), has transformed the educational landscape, providing unprecedented access to education for diverse populations [1]. However, despite these advancements, a significant challenge persists: a high dropout rate among students. Research indicates that approximately 78% of students enrolled in online courses do not complete their studies, primarily due to a lack of face-to-face interaction [2]. Another study states that the main reason for withdrawal from VLEs is the lack of student engagement, followed by the inability to locate materials or activities for assessments [3]. Therefore, student engagement is a key area of research, as low engagement negatively impacts final grades, knowledge retention, and dropout rates [4]. In VLEs, the more students engage in meaningful activities, the higher the probability that they will enjoy the course, perform well, and complete it [5].

This project aims to address the problem by applying machine learning (ML) models, comparing them, and selecting the best one as part of the Student Dropout Predictor (SDP). For this purpose, data from the Open University Learning Analytics Dataset (OULAD) is used [6]. The Open University, the largest in the UK, offers over a thousand online courses and provides full online degree programs [7].

This alarming statistic underscores the urgent need to develop effective strategies for early prediction and mitigation of student dropout, enabling timely support that can help learners stay engaged and successfully complete their courses.

1.1 ML Models for Predictions

ML, a part of artificial intelligence, allows computer programs to automatically find complex patterns in features taken from existing data. This helps in making smart decisions about new data [8]. ML algorithms are trained on sample data and later evaluated with unseen data [9]. This type of ML, where the algorithm learns a mapping function from input variables (features) to an output variable (label) using labelled training data, is known as Supervised Learning [10]. ML algorithms can provide instructors with real-time insights about students, enabling early interventions during the course [11]. ML is widely applied to develop predictive models from student data, handling both numerical and categorical variables to effectively model student behaviours and outcomes [12].

This project conducts a comparative evaluation of six ML models: Random Forest (RF), Logistic Regression (LR), K-Nearest Neighbors (KNN), LightGBM, Support Vector Machine (SVM), and Neural Networks (NNs). Each model is assessed on its ability to classify student outcomes into the discrete labels ‘Withdrawn’ (0), ‘Fail’ (1), or ‘Pass’ (2), with particular emphasis on accurately identifying the ‘Withdrawn’ class. For example, RF is known for its robustness and ability to handle large datasets with high dimensionality [13], while LightGBM is optimized for speed and efficiency, making it suitable for large-scale applications [14]. LR offers interpretable results [15], KNN captures local data patterns, SVM performs well in high-dimensional spaces [16], and NNs can model complex relationships [17]. Supervised learning algorithms can be divided into classifiers and regressors, where the previously ML models mentioned, when used as classifiers, serve to predict discrete class labels [18].

The primary evaluation metric will be recall for the ’Withdrawn’ class (0), as this reflects the model’s ability to correctly identify students who have dropped out. Maximizing recall for this minority class is critical to ensure timely support and intervention. Given the imbalanced nature of the classification problem, focusing on dropout recall helps prevent overlooking students who need immediate assistance [19]. The comparative analysis aims to identify which model achieves the highest recall for the 'Withdrawn' class, thereby determining the most effective approach for dropout detection.

1.2 Research Objectives

Using data from the Open University Learning Analytics Dataset (OULAD), the study will evaluate how well each model detects withdrawal cases, with a focus on maximizing recall for this class. Additionally, the project incorporates a 'Student Engagement' variable [20] and socio-economic factors such as the deprivation index to enrich the analysis of factors influencing student withdrawal.

Student engagement will be measured as a simple yes/no (binary) variable, following the method described by Mushtaq et al. [20]. This variable will be calculated using information such as scores on course assignments, whether the student passed or failed the course, and how active the student was in the virtual learning environment (VLE).
The specific objectives of this project are:

- To incorporate the student engagement feature as a meaningful predictor for the machine learning models.
- To identify which socio-demographic variables are meaningfully linked to student dropout.
- To select a predictive model capable of identifying students likely to withdraw at early stages of their course.

Earlier studies that used OULAD data, such as those by Tomasevic et al. [21] and Hussain et al. [20], have looked at student engagement and predicted student dropout. However, these studies did not fully examine which demographic variables showed stronger connections with dropout. The main reason for this project is to provide useful insights into the importance of demographic variables by improving the student dropout model with the addition of student engagement data. Additionally, this project will include the timing of assessments to help make predictions at an early stage of the course. Adding this analysis aims to offer another way to improve models that predict student dropout. For this project, the prediction results will be grouped into three classes: 'Withdrawal' (0), 'Fail' (1), or 'Pass' (2).

1.3 Significance and Contributions

The importance of this project extends beyond academic curiosity; it has practical implications for educators, administrators, and policymakers. By accurately predicting student dropout, institutions can implement timely interventions, tailor support services, and improve course designs, ultimately leading to higher retention rates and better educational outcomes [22]. Additionally, the insights gained from this research will benefit data scientists and educational technologists by providing a framework for developing predictive tools that can be applied across various educational contexts.

This project builds on existing literature that has explored student engagement and dropout prediction using machine learning techniques [21, 20]. However, it seeks to fill a critical gap by integrating demographic variables and engagement metrics into a comprehensive predictive model. The selected approach is significant as it not only addresses the immediate problem of dropout prediction but also contributes to the broader discourse on enhancing student engagement in online learning environments.

1.4 Challenges

Central challenges in this endeavour include ensuring the accuracy and reliability of the predictive models, addressing potential biases in the data, and effectively measuring student engagement at various stages of the learning process. These challenges are significant as they directly impact the validity of the predictions and the effectiveness of subsequent interventions. By tackling these issues, this project aims to provide valuable insights that can inform future research and practice in the field of online education.

In summary, this project focuses on performing and comparing six machine learning models to identify the best-performing model for detecting students who have withdrawn from online courses. By emphasizing recall for the 'Withdrawn' class and incorporating engagement and socio-economic factors, the study aims to provide actionable insights that support timely interventions and improve student retention in virtual learning environments.

Chapter 2

Technical Background

A compulsory chapter, roughly 20% of the total page-count

This chapter is intended to describe the technical basis on which execution of the project depends. The goal is to provide a detailed explanation of the specific problem at hand, and existing work that is relevant such as an existing algorithm that you use, alternative solutions proposed, supporting technologies, and relevant literature. The literature you include should cover appropriate peer-reviewed academic publications and books, and maybe also news and current-affairs articles published in reputable sources such as *The Economist* magazine or *The Financial Times* newspaper.

Your thesis should include complete set of references/bibliography: everything that you cite should be in there, in full. This means including the publisher's name for anything that is a book; the editors and title of books that a paper appears in as one item in a larger collection (e.g. proceedings volumes and/or thematic edited collections) and the relevant page-numbers; and so on. Put most simply, we require and expect your references to look like the references in a published peer-reviewed academic paper, so you can work out whether you still have work to do by comparing your references to the reference-list on the publications that you're working from.

Note there is a subtle difference from this and a full-blown literature *survey*. The latter might try to capture and organise (e.g., categorise somehow) *all* related work, potentially offering meta-analysis, whereas here the goal is simply to ensure that your thesis is self-contained. Put another way, after reading this chapter an intelligent non-expert reader with no prior knowledge of your project should have obtained enough background to understand what *you* have done (by reading subsequent sections), and then accurately assess your work. You might view an additional goal as giving the reader confidence that you are able to absorb, understand and clearly communicate highly technical material.

Just as there is no single ideal structure for an MSc thesis, there is no one correct name for this chapter. You could just call it *Background* if you wish, or *Technical Background*. Or if you feel you have a lot to say you could split this chapter into two: you might have your first three chapters being:

1. Introduction
2. Context
3. Related Work

But if you prefer to use a more concise structure, you could instead use:

1. Introduction: Contextual Background
2. Technical Background

The choice is yours.

The key thing is that by the end of these first two or three chapters you have told the reader everything they need to know so that they can understand the rest of your thesis. This is particularly important because at least one of the people who actually examines your thesis will be a UoB academic, one of our lecturers or professors, who you can assume knows almost nothing about the specifics of what work you have done, and who is given a copy of your thesis to mark. So what you write has to adequately explain things to that examiner. You can safely assume that whoever examines your thesis is numerate, intelligent, understands programming and data analytics etc, but you cannot safely assume that their specific individual expertise is a perfect match to the topic of your thesis (it's in that sense that the examiner is a *non-expert*). So, the person you write for, your *audience*, should be that undefined group of academics who might possibly examine your thesis: if you write for that audience and do a good job, your thesis should be understandable by a wide range of people, including potential employers and colleagues in the world of work.

Chapter 3

Execution

A topic-specific chapter, roughly 30% of the total page-count

This chapter is intended to describe what you did: the goal is to explain the main activity or activities, of any type, which constituted your work during the project. The content is highly topic-specific. For some projects it will make sense to split the content into two main sections, or maybe even into two separate chapters: one will discuss the design of something, including any rationale or decisions made, and the other will discuss how this design was realised via some form of implementation. You could instead give this chapter the title “Design and Implementation”; or you might split this content into two chapters, one titled “Design” and the other “Implementation” .

Note that it is common to include evidence of “best practice” project management (e.g., use of version control, choice of programming language and so on). Rather than simply a rote list, make sure any such content is useful and/or informative in some way: for example, if there was a decision to be made then explain the trade-offs and implications involved.

3.1 Example Section

This is an example section; the following content is auto-generated dummy text. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus foo

Figure 3.1: This is an example figure.

foo	bar	baz
0	0	0
1	1	1
⋮	⋮	⋮
9	9	9

Table 3.1: This is an example table.

convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

3.1.1 Example Sub-section

This is an example sub-section; the following content is auto-generated dummy text. Notice the examples in Figure 3.1, Table 3.1, Algorithm 3.1 and Listing 3.1. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat

for $i = 0$ **upto** n **do**
  $t_i \leftarrow 0$
end

Algorithm 3.1: This is an example algorithm.

```
for( i = 0; i < n; i++ ) {
  t[ i ] = 0;
}
```

Listing 3.1: This is an example listing.

magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Example Sub-sub-section

This is an example sub-sub-section; the following content is auto-generated dummy text. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Example paragraph. This is an example paragraph; note the trailing full-stop in the title, which is intended to ensure it does not run into the text.

Chapter 4

Critical Evaluation

A topic-specific chapter, roughly 30% of the total page-count

This chapter is intended to evaluate what you did. The content is highly topic-specific, but for many projects will have flavours of the following:

1. functional testing, including analysis and explanation of failure cases,
2. behavioural testing, often including analysis of any results that draw some form of conclusion wrt. the aims and objectives, and
3. evaluation of options and decisions within the project, and/or a comparison with alternatives.

This chapter often acts to differentiate project quality: even if the work completed is of a high technical quality, critical yet objective evaluation and comparison of the outcomes is crucial. In essence, the reader wants to learn something, so the worst examples amount to simple statements of fact (e.g., “graph X shows the result is Y”); the best examples are analytical and exploratory (e.g., “graph X shows the result is Y, which means Z; this contradicts [1], which may be because I use a different assumption”). As such, both positive *and* negative outcomes are valid *if* presented in a suitable manner.

Chapter 5

Conclusion

A compulsory chapter, roughly 10% of the total page-count

The concluding chapter(s) of a dissertation are often underutilized because they're too often left too close to the deadline: it is important to allocate enough time and attention to closing off the story, the narrative, of your thesis.

Again, there is no single correct way of closing a thesis.

One good way of doing this is to have a single chapter consisting of three parts:

1. (Re)summarise the main contributions and achievements, in essence summing up the content.
2. Clearly state the current project status (e.g., "X is working, Y is not") and evaluate what has been achieved with respect to the initial aims and objectives (e.g., "I completed aim X outlined previously, the evidence for this is within Chapter Y"). There is no problem including aims which were not completed, but it is important to evaluate and/or justify why this is the case.
3. Outline any open problems or future plans. Rather than treat this only as an exercise in what you *could* have done given more time, try to focus on any unexplored options or interesting outcomes (e.g., "my experiment for X gave counter-intuitive results, this could be because Y and would form an interesting area for further study" or "users found feature Z of my software difficult to use, which is obvious in hindsight but not during at design stage; to resolve this, I could clearly apply the technique of Bloggs *et al.*.

Alternatively, you might want to divide this content into two chapters: a penultimate chapter with a title such as "Further Work" and then a final chapter "Conclusions". Again, there is no hard and fast rule, we trust you to make the right decision.

And this, the final paragraph of this thesis template, is just a bunch of citations, added to show how to generate a BibTeX bibliography. Sources that have been randomly chosen to be cited here include:

Bibliography

[1] I. Maiz Olazabalaga, C. Castaño Garrido, and U. Garay Ruiz. Research on MOOCs: Trends and Methodologies. *Porta Linguarum, Monográfico I*, pages 87–98, 2016.

[2] O. Simpson. '22% -can we do better?'-The CWP Retention Literature Review. *ResearchGate*, 2010.

[3] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal, J. Vaclavek, and A. Wolff. OU Analyse: analysing at-risk students at The Open University. *Learning Analytics Review*, LAK15-1:1–16, 2015.

[4] A. Staikopoulos, I. O'Keeffe, B. Yousuf, O. Conlan, E. Walsh, and V. Wade. Enhancing Student Engagement through Personalized Motivations. In *2015 IEEE 15th International Conference on Advanced Learning Technologies*, pages 340–344, 2015.

[5] Y. Jung and J. Lee. Learning Engagement and Persistence in Massive Open Online Courses (MOOCs). *Computers & Education*, 122:9–22, 2018.

[6] J. Kuzilek, M. Hlosta, and Z. Zdrahal. Open University Learning Analytics dataset. *Scientific Data*, 4(1):170171, 2017.

[7] M. Hlosta, D. Herrmannova, L. Vachova, J. Kuzilek, Z. Zdrahal, and A. Wolff. Modelling student online behaviour in a virtual learning environment, 2018.

[8] J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. The MIT Press, first edition, 1992.

[9] S. Kotsiantis, C. Pierrakeas, and P. Pintelas. Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. *Applied Artificial Intelligence*, 18(5):411–426, 2004.

[10] K. P. Murphy. *Machine learning: A probabilistic perspective*. MIT Press, 2012.

[11] S. Kai, J. M. L. Andres, L. Paquette, R. S. Baker, K. Molnar, H. Watkins, and M. Moore. Predicting student retention from behavior in an online orientation course. In *Proceedings of the 10th International Conference on Educational Data Mining (EDM 2017)*, pages 250–255, 2017.

[12] R. S. Baker and P. S. Inventado. Educational Data Mining and Learning Analytics. In J. A. Larusson and B. White, editors, *Learning Analytics: From Research to Practice*, pages 61–75. Springer, 2014.

[13] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.

[14] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

[15] F. E. Harrell. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer International Publishing, second edition, 2015.

[16] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[18] A. Géron. *Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Inc., 2 edition, 2019.

[19] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.

[20] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. *Computational Intelligence and Neuroscience*, 2018(1):6347186, 2018.

[21] N. Tomasevic, N. Gvozdenovic, and S. Vranes. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143:103676, 2020.

[22] E. R. Kahu. Framing student engagement in higher education. *Studies in Higher Education*, 38(5):758–773, 2013.

Appendix A

An Example Appendix

Content which is not central to, but may enhance the dissertation can be included in one or more appendices; examples include, but are not limited to

- lengthy mathematical proofs, numerical or graphical results which are summarised in the main body,
- sample or example calculations, and
- results of user studies or questionnaires.

Note that in line with most research conferences, the examiners are not obliged to read such appendices.