The University of Sydney Faculty of Engineering School of Electrical and Information Engineering Year

A thesis submitted in partial fulfilment of a Bachelor of Software Engineering Degree

Student Name	Sophia Angela Jennifer Polito
SID	460353465
Unit of Study Code and Name	ELEC4713
Supervisor	Dr. Irena Koprinska Dr. Bryn Jeffries
Title	Machine learning methods for exploring student's learning style and procrastination behaviours in an online programming course

The University of Sydney

SCHOOL OF ELECTRICAL AND INFORMATION ENGINEERING

PROJECT CLEARANCE FORM

Unit of Study Code and Name ELEC4713: Thesis B

This is to certify that my student

Student Name Sophia Polito

SID 460353465

Has:

- **Returned all books and reference material;**
- **Returned all equipment and keys; and**
- **Tidied their work place.**

SUEIE Academic supervisor:

Signature: **Date:**

Name:

External supervisor (if applicable):

Signature: Irena Koprinska Digitally signed by Irena Koprinska Date: 2021.11.05 12:40:12 +11'00' **Date:** 5/11/2021

Name: A/Prof Irena Koprinska (School of Computer Science)

Statement of achievements

The thesis contributions are outlined in Section 1.5 Contributions and are as follows:

1. Learning style features were defined for a computer programming course and learning style was measured based on behavioural features and procrastination features. Behavioural features are based on Felder-Silverman's theory [9], and the procrastination features are based on Cerezo et al.'s work [11]. Using these features, the student vectors were clustered to investigate similarities in student's behavioural or procrastination tendencies and compare the two approaches' cluster meaningfulness and ability to identify different groups of high and low performing students. Overall, by comparing the behavioural and procrastination clustering approaches, our thesis contributes a meaningful insight into which approach best suits the nature of the NCSS course data and can therefore be best utilised by educators to identify struggling students.
2. A novel investigation into the overlap between the behavioural and procrastination groups was undertaken to view the procrastination trends within each behavioural cluster of students. Previous research would cluster students using either behavioural or procrastination features, however this thesis highlights the complexities of student learning styles as encompassing both procrastination and behavioural tendencies.
3. Gender demographic profiling of student in each behavioural and procrastination cluster was conducted for in-depth analysis on the students within each learning style. This is a significant contribution to the research which aims to support females in programming fields by supporting their learning and performance in online programming courses.
4. Changes in student learning style throughout the course are investigated using Markov chains which differs from existing literature where changes in student learning style is determined through clustering the number of times a student belongs to each learning style cluster. Markov chains can instead provide a more human-readable visualisation for educators to determine how student learning styles evolve throughout the course. Changes in student learning behaviour can be an important indicator for a number of course outcomes such as course difficulty, as well as indicate student disengagement from the course.
5. Markov chain analysis indicated when high risk students were most susceptible to course disengagement. Therefore, the results of this thesis can inform educators on when at-risk students require early intervention methods prior to course drop-out, allowing educators to re-engage students with the course material.
6. We demonstrate the application of these methods using a large dataset of 10,558 students and 52,790 vectors, where one vector was created for each student for each of the 5 weeks

of the Grok NCSS beginners course duration. The initial 52,790 vectors were filtered to 15,733 behaviour vectors and 19,724 procrastination vectors after removing vectors with missing values. For the Markov chains the initial 52,700 vector dataset was filtered to 44,160 vectors after removing inactive students with no activity for any week of the course.

7. A conference paper on this research was submitted to the 2022 Australasian Computing Education Conference (ACE) and this paper is attached in Appendix A.

Abstract

As programming courses are increasingly being taught online, educators are faced with the problem of assisting struggling students remotely. Existing online programming courses use automatic testcase grading systems which feed a set of inputs into the student's program, and then compare the student's program output to the expected output. In addition to testcases, data driven methods such a data mining present an opportunity for educators to seek further insight on their student's performance and learning style. Whilst educators can view student scores through automatic testcase grading of programming submissions, additional student learning style indicators can inform educators on how students prefer to learn, how different groups of students perform, and how these students can respond to feedback to improve their learning. Through analysing student's learning behaviour, teachers can better identify at-risk students and take remedial action to improve the student's online learning experience and course performance.

In this thesis, we investigate different types of student behaviour through data mining of the Grok National Computer Science School (NCSS) Python programming course. Four main research questions are explored, the first objective is concerned with determining trends in student learning style, which will be analysed using Felder-Silverman's behavioural categories and procrastination features, using K-means clustering. Secondly, the student clusters will then be investigated to determine how a student's learning style impacts a student's performance and course mark. Thirdly, demographic profiling of the clusters will be undertaken to consider if higher or lower performing student clusters have different gender distributions. The final objective seeks to analyse how student submission behaviour changes over the duration of the course. This is explored using Markov chains to determine how students move between clusters throughout the 5 weeks of the course. A number of contributions in this thesis have been made, and consist of: defining behavioural and procrastination features based on existing literature, investigating the overlap between the behavioural and procrastination groups which is a novel exploration of student learning tendencies, gendered demographic profiling of low and high performing clusters to better support females in programming courses, utilising Markov chains to explore transitions in student behaviour throughout the course, and applying the methods to an initial dataset of 10,588 students and 52,790 vectors. Overall, this thesis classifies students based on their learning style and provides insight into how these learning tendencies change throughout the course duration. As a result, this thesis is significant for educators who want to identify at-risk students and know when these students are most susceptible to disengaging from the course, to allow for better remedial action and student support.

Acknowledgements

I would like to acknowledge my supervisors Dr. Irena Koprinska, and Dr. Bryn Jeffries, for their guidance over the last year. Their consistent feedback, encouragement, and sharing of their technical knowledge is what made this thesis possible. I could not have asked for more supportive, reliable, and knowledgeable supervisors throughout this thesis.

I would also like to thank my family and grandparents for supporting me throughout my university degree. Especially my Nonna Surace who always encouraged me to pursue my education, I wish you could be here to see me graduate, and you will be missed.

Table of contents

List of figures

List of tables

Chapter 1 Introduction

1.1 Thesis introduction

The introduction chapter aims to provide background context on online programming courses and to outline the significance of the thesis research in the online programming education field. The limitations of existing online programming course feedback mechanisms will be discussed, as well as the opportunity for using data driven methods to categorise students based on their learning tendencies to identify at-risk students. This section will conclude with an overview of the Grok learning website which will be used as a data source for this thesis, followed by the research objectives and contribution.

1.2 Background

1.2.1 Online courses

Programming is becoming an increasingly important skill in today's technological society and online courses can provide students with programming course content as well as automatic feedback on programming submissions through auto-graded testcases. Whilst the auto-grading testcases allow for efficient grading, they are not always as effective at providing tailored feedback to students. Additionally, teachers can struggle to provide effective feedback to students due to the high number of enrolments in online courses and the often time-consuming analysis needed to manually assess each student's program and behaviour. Therefore, there is an opportunity to use data driven methods to provide insights into student's learning behaviour and performance, to better identify at-risk students who may need additional supports throughout online programming courses.

1.2.2 Testcases and feedback methods in online courses

Existing feedback methods in online programming courses generally consist of auto-graded testcase feedback. When a student submits their code, it is tested against several inputs and corresponding expected outputs. The testcases will then inform a student on which cases were passed and which testcases the code failed. In the case of a failed testcase the student is often provided with a short description of the testcase or a comparison of the expected output with their own output. Online programming systems can use different variations of testcase feedback with the main variation being the level of visibility a student has over what is being tested in their code and which testcases they failed.

A limitation of testcases is they can only provide feedback on the outputs of the code. Whilst in some instances an incorrect program output can give insight into the student's programming logic, it can be difficult to ascertain mistakes in the student's logic from output alone. This can be a source of frustrations for students when attempting to improve their programming submission, and for educators when assisting student's learning as code output alone does not always indicate the cause for the testcase failure. Therefore, the lack of feedback on code logic is a limitation of testcases in allowing student to improve their submissions.

In traditional classrooms, feedback provided for a student's program logic would be completed by the educators manually. This feedback could be in the form of informal feedback during classroom sessions or formal feedback during the assignment marking process. In-person classes also allow instructors to identify struggling students and provide the appropriate guidance and feedback for those students. Therefore, there is an opportunity to simulate in-class feedback and better identification of student needs using data and machine learning techniques.

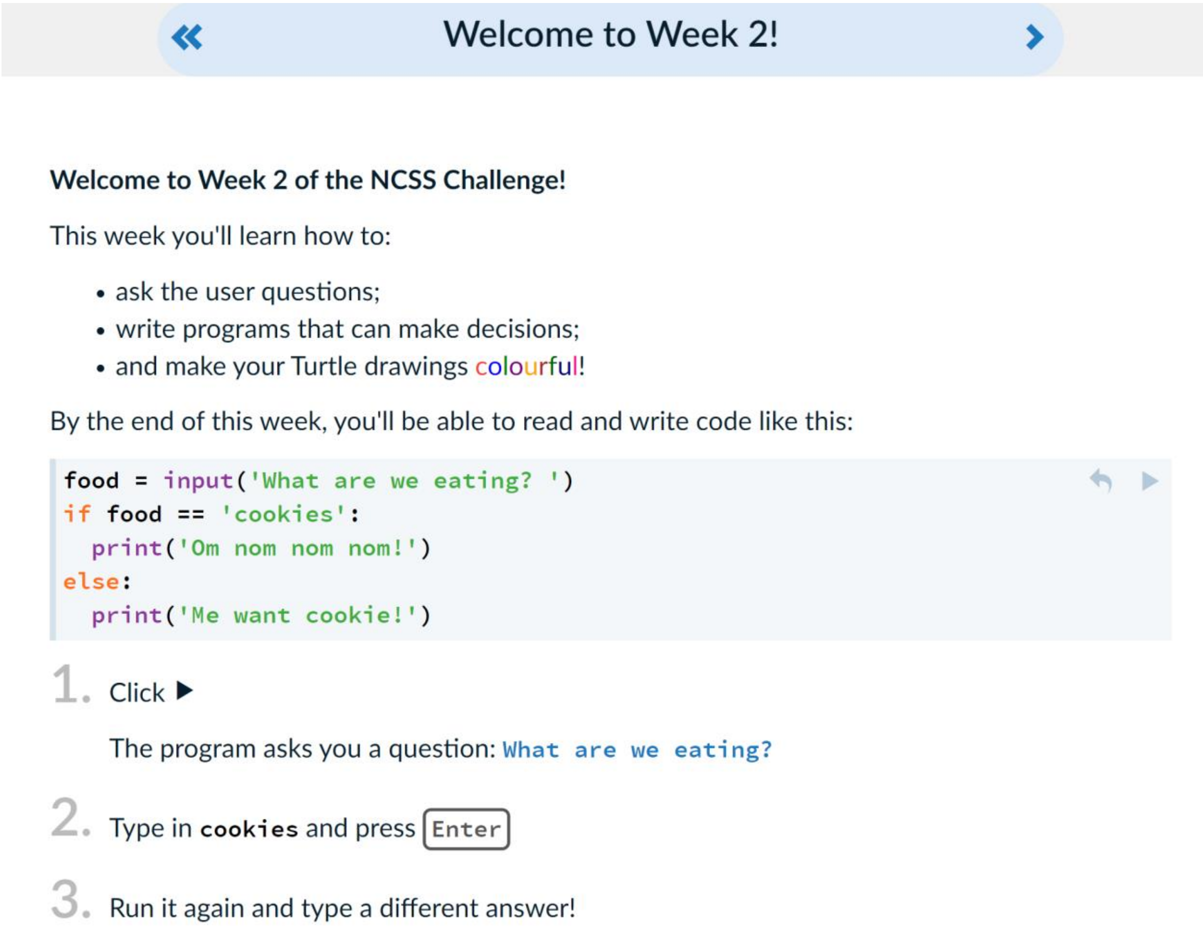
1.2.3 Grok learning platform

Grok Learning is an online programming website which provides programming courses for primary to high school aged students. Grok offers the National Computer Science School (NCSS) course which is open to Australian school students nationwide for different levels of Python programming over a 5-week period, in two different competition sessions each year. The NCSS course is offered at a beginners Blockly level, beginners Python, intermediate Python and advanced Python course. The Blockly course teaches students a visual click-and-drag language based off of Python. The courses provide content to teach students Python programming concepts, as well as scored programming problems which contribute to a student's ranking in the course. Students can be enrolled through their school curriculum or through their own initiative. The data for this thesis was sourced from the Grok NCSS 2019 beginners Python course.

1.2.3.1 Content slides

Grok teaches programming concepts to students on their content slides. Content slides provide explanations on topics such as variable types, input and output, logic flow, if statements and loops. Slides also contain example code, programming questions for students to answer, and steps to help guide students through the code examples as seen in Figure 1.1. Questions on content slides do not contribute to a student's overall course score and students are allowed unlimited attempts at questions on content slides without penalty. All content for the week is non-compulsory and students can progress to problems without viewing or interacting with the slides. There are also no guards to prevent students from exploring the content out of order for the week, therefore students have the freedom to progress through the course to their preferred learning style.

FIGURE 1.1 USER INTERFACE OF A GROK NCSS COURSE CONTENT SLIDE



1.2.3.2 Problems and testcases

The 5-week Grok NCSS beginners Python course consists of 8 graded problems each week, with a total of 40 problems in the course. The problems test the programming skills taught in the content slides provided earlier in the course. The release of each week's content and problems is staggered, with students spending a week on the content and problems before the next set of content and problems are released. Grok Learning implements a testcase based system of grading programming submissions where students can see which testcases they passed or failed. In the event of a failed testcase, a description of the error is displayed and for some problems, the testcases also provide suggestions on how to improve the submission. Students can also message volunteer tutors in the chat features for assistance. An example problem is shown in Figure 1.2 while the testcase interface is shown in Figure 1.3 and Figure 1.4.

FIGURE 1.2 USER INTERFACE OF A GROK NCSS PROBLEM

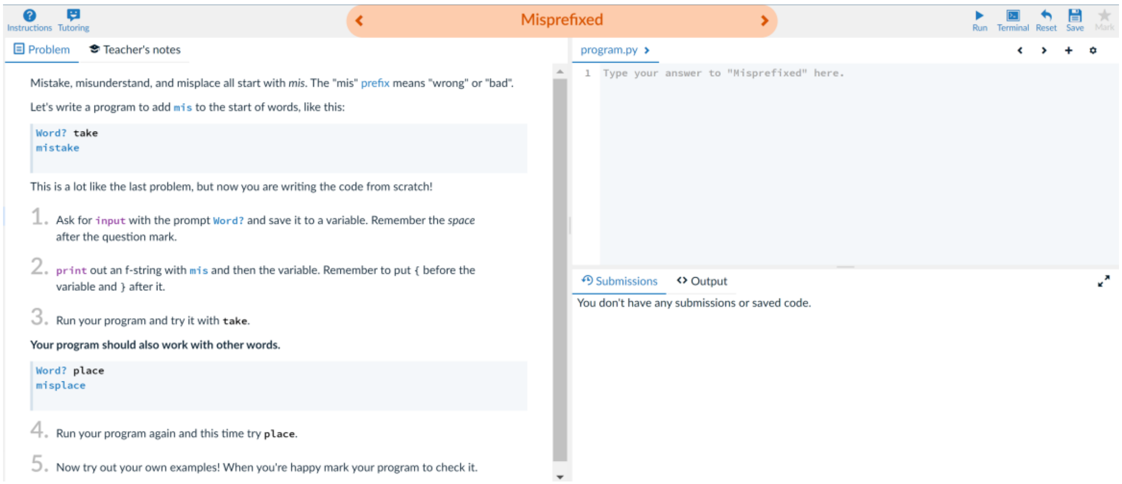


FIGURE 1.3 USER INTERFACE OF TESTCASES FOR A SUCCESSFULLY COMPLETED PROBLEM

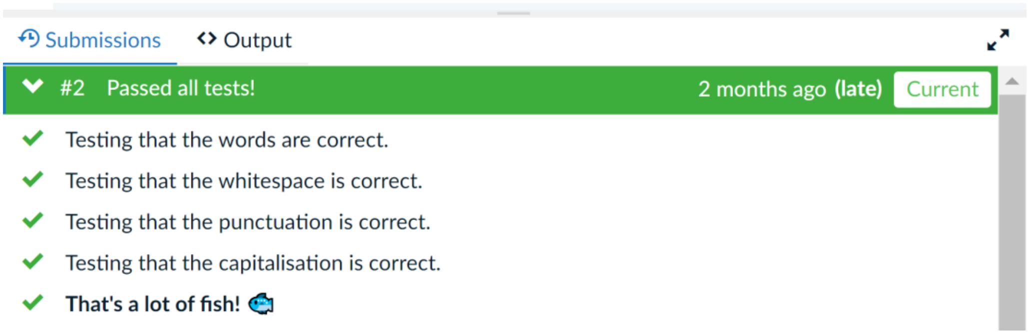
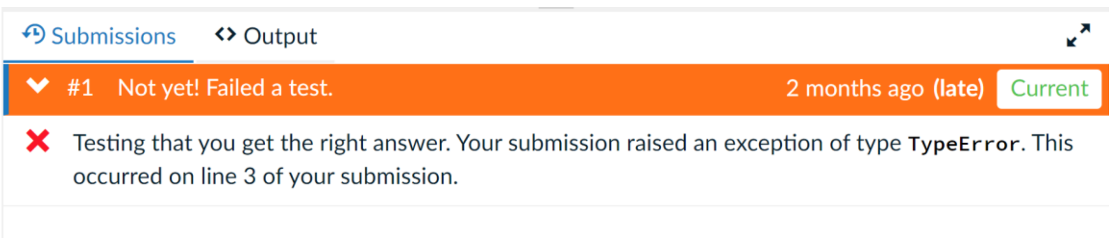


FIGURE 1.4 USER INTERFACE OF TESTCASES FOR AN UNSUCCESSFULLY COMPLETED PROBLEM



1.2.3.3 Score

Marks are awarded for correctly answered problems that pass all testcases. There are 8 scored problems for each of the 5 weeks of the course and these problems are marked by auto-grading testcases. The highest possible score in the course is 400 with a maximum of 10 possible marks awarded for each correct problem. Marks are deduced after every 5 incorrect submissions a student makes. These deductions are capped at 5 marks, so a student with a very high number of submission attempts will still achieve 5 marks once they pass all of the testcases for a question. If a student does not attempt a question or is unable to pass all testcases they receive a 0 score for that problem. A graded submission is different to a terminal run, where a terminal run allows a student to infinitely run their program for a problem without feedback from the testcases and without penalty.

1.3 Definitions

Throughout the thesis, the following terms are used: learning style, learning behaviours and procrastination tendencies. Learning styles is used as a broad term to encapsulate both learning behaviours and procrastination tendencies. Behavioural learning is described by the Felder-Silverman model (FSLSM) which outlines different categories on how students receive, process and interact with learning content. Conversely, procrastination tendencies are defined based on the amount of time or delay a student takes to complete a task, as well as how many elements of the course a student completes. Both time and effort related measures are taken into consideration for procrastination to provide insight on how students spent their time in the course. For example, a low amount of time spent to complete a problem could be due to a student completing the problem quickly, indicated by a high problem completion rate, or because the students barely attempted the problem, indicated by a zero or low problem completion rate.

1.4 Thesis objectives

Overall, this thesis aims to explore the learning styles of students in the Grok NCSS beginners Python course to allow educators to identify students at risk of course disengagement and poor performance, using data driven methods. Grok learning data will be used to identify similarities in students, both within the student's learning style, student's scores, and non-behavioural attributes such as their gender. The changes in student learning style will be also examined over the course of the competition duration.

The thesis aims can be summarised in the following research objectives:

1. Investigate similarities in student learning style, which will be analysed in terms of learning behaviours and procrastination tendencies.
2. Examine how learning behaviours and levels of procrastination influence student's overall course score.
3. Determine if non-behavioural attributes of a student such as gender, impacts their learning style and score.
4. Investigate how student learning style changes over time and impacts drop-out rates.

1.5 Contributions

The research contributions are as follows:

1. Learning style features were defined for a computer programming course and the learning style was measured based on behavioural features and procrastination features. Behavioural features are based on Felder-Silverman's [9] theory, and the procrastination features are based on Cerezo et al.'s [11] work. Using these features, the student vectors were clustered to investigate similarities in student's behavioural or procrastination tendencies and compare the two approaches' cluster meaningfulness and ability to identify different groups of high and low performing students. This contribution will be outlined in research objective 1 and 2. Overall, by comparing the behavioural and procrastination clustering approaches, our research contributes a meaningful insight into which approach best suits the nature of the NCSS course data and can therefore be best utilised by educators to identify struggling students.
2. A novel investigation into the overlap between the behavioural and procrastination groups was undertaken to view the procrastination trends within each behavioural cluster of students. Previous research would cluster students using either behavioural or procrastination features, however this thesis highlights the complexities of student learning styles as encompassing both procrastination and behavioural tendencies.
3. Gender demographic profiling of student learning styles is conducted for in-depth analysis on the students within each learning style group in research objective 3. This is a significant contribution to the research which aims to support females in programming fields by supporting their learning and performance in online programming courses.
4. Changes in student learning style throughout the course are investigated using Markov chains which differs from existing literature where changes in student learning style is determined through clustering the number of times a student belongs to each learning style cluster. Markov chains research in objective 4, instead provides a more human-readable visualisation for educators to determine how student learning styles evolve throughout the course. Changes in student learning behaviour can be an important indicator for a number of course outcomes such as course difficulty, as well as indicate student disengagement from the course.
5. Markov chain analysis indicated when high risk students were most susceptible to course disengagement. Therefore, the objective 4 results of this thesis can inform educators on when at-risk students require early intervention methods prior to course drop-out, allowing educators to re-engage students with the course material.

6. We demonstrate the application of these methods using a large dataset of 10,558 students and 52,790 vectors, where one vector was created for each student for each of the 5 weeks of the Grok NCSS beginners course duration. The initial 52,790 vectors were filtered to 15,733 FSLSM behaviour vectors and 19,724 procrastination vectors used for the final clustering after removing vectors with any missing values. For the Markov chains the initial 52,700 vector dataset was filtered to 44,160 vectors after removing inactive students who had no activity for any week of the course.
7. A conference paper on this research was submitted to the 2022 Australasian Computing Education Conference (ACE) and this paper is attached in Appendix A.

Chapter 2 Literature review

2.1 Machine learning methods

Machine Learning is a branch of Artificial Intelligence where models learn from past data to make predictions on new data examples [1]. There are two types of machine learning, supervised learning and unsupervised learning. Supervised learning models use training data which contains classifications of each data example in the dataset. Conversely, unsupervised learning does not require the data to have a classification [1]. Since the NCSS Grok learning data used for this thesis does not contain classifications, unsupervised learning methods will be considered for this thesis.

2.2 Unsupervised learning - clustering

The main unsupervised algorithms are clustering, anomaly detection, neural networks and latent variable models [1]. Additionally, Principle Component Analysis can be used for feature selection in unsupervised data.

Cluster analysis algorithms group together similar data points to produce clusters which have data that is similar to other data points in the same cluster, but are different from the data in the other clusters [1]. A distance function is used to create clusters with high cohesion and high separation, where cohesion refers to the similarity within clusters, and separation refers to the similarity between different clusters [2].

The different types of clusterings (entire collections of clusters) include hierarchical, partitional, exclusive, overlapping, fuzzy, complete and partial [2]. Hierarchical clusters are nested and have subclusters organised in a tree format whilst partitional clusters divide the data into *k* unique clusters where *k* is provided as an input to the algorithm. Exclusive clusters have each object in one single cluster, whilst overlapping clusters can have data belonging to more than one group at the same time. For fuzzy clustering, each data point belongs to every cluster but has a membership weight to indicate how much the data belongs to the cluster. Finally, complete clusters have every data point in a cluster, whilst partial clustering has some data not included in any cluster.

Clustering techniques include K-means which groups data based their distance to centroids. The number of clusters is provided as an input and the output is a *k* number of partitional clusters [2]. Alternative clustering techniques include agglomerative hierarchical clustering which begins with each point as a single cluster and iteratively merges the closest clusters until only one reminds. DBSCAN is density-based clustering where points low in density are classified as noise and omitted whilst the other data points are assigned to number of clusters as determined by the algorithm [2].

2.3 Educational data mining

Educational data mining (EDM) aims to 'analyse educational data to develop models to improve learning experiences and institutional effectiveness' [3]. There are a number of objectives in EDM which include the creation or improvement of a student model, predictive modelling, recommendation generation, student behavioural analysis and course maintenance. Ugalde and Radhakrishnan [4] outline methods used in EDM which include prediction, clustering, relationship mining, discovery and visualisation of data.

Generally, there are three main stakeholders in EDM which consist of teachers, students and management [4]. Teachers develop and distribute course material as well as marks assessments, whilst students undertake the course, and management governs the courses. However, in the context of online learning, stakeholders can also include the platforms which distribute learning material and in some cases teachers may have less of an importance as a stakeholder.

2.4 Importance of feedback

Different types of feedback have been explored by Kumar's and Stracke's [5] research which aimed to identify the types of feedback given to students on their Masters and doctoral thesis submissions. They identified and coded two main types of marker's annotations: assessment and feedback. Assessment can be defined as a measure of competence and comments on if a student has met the expected outcomes of the task. Conversely, feedback is defined as closing the 'gap between current and desired performance'. Within the feedback category of marker's annotations, Kumar and Stracke referred to three different types of feedback: 'feed up' which identifies what a learner should be achieving and is usually provided prior to students starting a task, 'feed back' referring to what progress has been made to achieve a goal, and 'feed forward' to identify what guidance a learner should be given next based on past performance. 'Feed forward' feedback was seen as the most useful feedback for students because of its specific nature which provides recommendations to students on what to work on or poses questions to prompt student reflection.

2.4.1 Summary

A limitation of Kumar's and Strake's study is their findings are qualitative and no quantitative results are presented regarding the impact of different feedback methods on student improvement. However, their research highlights the importance of combining assessment comments, which identify if students are meeting goals in conjunction with 'feed forward' feedback to allow students to bridge the gap between their existing knowledge and the expected outcomes.

This study could prove important in evaluating the current feedback methods on Grok Learning and the impact of feedback on student's coding progress. The existing Grok testcases mirror the 'assessment' style grading discussed in this study whilst the Grok recommendations provided testcases are run incorporates 'feed forward' feedback. Although Grok implements 'feed forward' feedback, students may respond to this feedback in different ways based on their learning style and confidence in the course material. Hence there is an opportunity to investigate how students are able to respond to feedback to close gaps in their existing knowledge.

2.5 Psychological theory of student behaviour

Student behaviour can be defined in a number of ways and psychology literature has provided insight into different influences on learning behaviour such as motivation theory and learning style theory.

2.5.1 Motivation theory

Motivation has shown to be a factor that can influence a student's studying style and behaviour both for in-person and online classes. Biggs [6] created different classifications of students based on their motivation for study: utilising students, internalising students and achieving students. Utilising types of students are characterised as studying for the purpose of avoiding failure, resulting in these students becoming syllabus bound and doing as little work as possible. Internalising students generally have intrinsic motivation and use learning to actualise their own purpose or interests. Achieving students see learning as competitive and are intrinsically motivated to achieve high grades. It was found that students who can align their own learning goals with those prescribed by the courses generally perform better than those who do not.

Study motivation can result in different studying techniques such as surface study and deep study. Surface studiers 'reproduce signs of learning' [6] rather than focus on understanding the content and shows similarities with anxious prone Utilizing students. Deep study which aims to master content is linked to Internalising student motivation.

Using a Study Process Questionnaire (SPQ) on 60 undergraduate students, Biggs [5] compared student's study orientation to the quality of short response answers to two questions. Results showed that Achieving orientations had low response complexity but high recall of facts. This recall was only temporary lasting on average a week. The Internalising orientations lead to more complex response levels whilst Utilising students had high retention of facts and low complexity of response.

Conversely, Li and Tsai [7] analysed the logs of a Learning Management System (LMS) for a computer science course to determine the effects of motivation and learning performance. Students were clustered using K-Means into 'consistent use' students (cluster 1) who used all the course material, 'less use' students (cluster 2) who used minimal LMS resources and 'slide intensive use' students (cluster 3) who used the lecture slides primarily. Cluster 1 students spent the most time viewing video lectures, assignment content and notifications whilst cluster 3 primarily viewed lecture slides and cluster 2 spent the least amount of time viewing any materials. Results showed the highest student performance on homework tasks were ‘consistent use’ students, followed by ‘slide intensive’ students then by ‘less use’ students. However, the difference between consistent use students and slide intensive students in the final exam was minimal. Li and Tsai then compared the students in the clusters for their motivational strategies as determined by a Motivated Strategies for Learning Questionnaire (MSLQ) which asked questions on intrinsic and extrinsic goal orientation, task value, control belief of learning, self-efficacy and test anxiety. Pairwise Mann-Whitney-U-Tests showed ‘consistent use’ students had higher intrinsic goal orientation, self-efficacy and task value. Consistent use students also showed higher self-efficacy and intrinsic goal orientation compared to slide intensive use students.

2.5.1.1 Summary

Whilst both Bigg’s [6] research as well as Li’s and Tsai’s [7] research demonstrated that student motivation can influence learning performance, literature has also shown that motivation alone is complex and sensitive to situational context such as deadlines and course progression [8]. Therefore, motivation alone may not be a reliable indicator of student behaviour and will not be further explored in this thesis.

2.5.2 Learning style behavioural theory

Learning style can be defined through behavioural measures in education learning theory such as the Felder-Silverman Learning Style Model (FSLSM) [9]. FSLSM is a scale-based model which characterises engineering students based on the following learning styles: active/reflective, sensing/intuitive, sequential/global and visual/verbal. Active learners are defined as being collaborative and hands on whilst reflective learners prefer to work individually and have an introspective learning process. Sensing learners prefer practical examples whereas intuitive learners prefer theories and conceptual learning. Visual learners prefer diagrams and pictures whilst verbal learners prefer written or verbal content. Sequential learners generally think linearly and in-order whilst global learners can take large leaps in learning and think holistically. Visual learners prefer to learn from images and graphics, whilst verbal learners prefer audio and written communication.

FSLSM features were utilised in Graf’s et al. [10] research to determine insights on learner’s navigational behaviour in an online Object Orientated Modelling (OOM) course. This study used the Index of Learner Styles (ILS) questionnaire where students answer questions to self-report their behavioural learning styles. Visual and verbal learners FSLSM categories were omitted from this study.

Graf et al. [10] proved a number of assumptions on the types of learners and how they use an online learning system. Active students utilised the discussion board to ask and answer questions and they made regular coding submissions since they prefer to try things first-hand. They were also shown to prefer reading summaries of information rather than detailed chapters. Conversely, reflective learners were shown to use the discussion board as a second source of information, made less submissions and were less likely to jump between subjects or problems. Reflective learners were more inclined to read chapters of information and visited self-assessment tests after reviewing question feedback before progressing to new content or questions.

Sensing learners used the forum more than intuitive learners primarily to clarify details of the assessments. The also navigated more to quiz pages since they generally prefer to be well prepared when applying theory to new practical problems. They showed higher rates of navigation to access quizzes particularly on overview pages and reviewed feedback for self-assessments more frequently. Intuitive learners were shown to go from one problem or test to another without necessarily completing the previous one as they prefer not to focus on details.

Global learners spent more time on overview content and preferred conclusions to aid them in obtaining a high-level understanding. Sequential learners in comparison progressed through the course in order and returned to the self-assessment pages after reviewing content.

2.5.2.1 Summary

Graf et al. [10] were able to link student behaviour with the FSLSM learning theory and gain insight into how different types of students navigated through their online learning website. The methods in this study are replicable with online learning systems which log student navigation and interaction with the website. Therefore, these FSLSM learning theory classifications would prove useful for clustering navigation data such as the data to be used in this thesis.

However, a limitation is the study was able to link student behaviour and learning theory through ILS surveys which allowed learners to answer questions to identify themselves as a particular learning style classification. Whilst this allows for comparison of student's actual behaviour against their self-reported learning style, this type of study relies on questionnaires for students which may not be possible for historical data or for a large number of students.

2.5.3 Procrastination behaviours

Student's procrastination tendencies are also used to define student learning styles in online courses. Cerezo et al. [11] measured student behaviour in terms of procrastination levels using six variables to generate four clusters using K-means. Clusters were based on student's level of procrastination and how much time was spent on quizzes compared to lecture content. Variables used for clustering included time spent on quizzes, time spent viewing theoretical content, time spent on forums, number of words written on forums, number of other relevant actions and number of days students took to hand in tasks. Their results showed students who spent more time on the quizzes performed better than students who did not and that final exam score was not related to time spent on viewing theoretical content or time spent on lecture slides.

Hooshyar, et al. [12] developed a novel algorithm called PPP to predict student procrastination behaviour to an accuracy of 96%. Procrastination was defined by spare time and inactive time. Spare time is the time from when a student submits an assignment until the assignment is due, and inactive time is the time between when an assignment is open and a student views it for the first time. Other features such as student submission date and deadline also indicated any late submissions made by students. Using K-means, clusters were then categorised into non-procrastinating students, potential procrastinators and procrastinators. Statistical analysis showed there was a positive correlation between spare time and assignment scores and a negative correlation between assignment score and inactive time.

2.5.3.1 Summary

Procrastination related features are shown to have high accuracy in predicting student learning style and performance in online courses. In particular, it is important to note that procrastination features in these studies included not only features such as late submissions, but also general behavioural features of students observed before a due date. Therefore, deadline related procrastination behaviours alone may not be sufficient to adequately cluster students and procrastination features should be considered within the context of other behavioural features.

2.6 Attribute selection

Student learning style can be defined in numerous ways, and this section aims to explore the existing literature on attribute selection as well as highlighting attributes that existing studies have found to be of notable importance in data mining of student behaviour.

A combination of time and submission related features have been used by McBroom et al. [13] to cluster students in an online programming course. Time based features include the time taken to complete a task as well as percentage of attempts made early, on time or late. Submission attributes consisted of the number of attempts, number of compilation errors, first mark and last mark. Behavioural changes were also considered and were measured based on how long each student's submissions belonged in the behavioural clusters over the semester. This allowed for a measure of student behavioural changes over time and identifies how student's approach to the course changed throughout the semester.

Conversely Chow [14] clusters students based on a broad definition of behaviour including the type of edits, improvement, consistency and submission behaviour. The type of edit attributes includes number of major or minor edits, number of structural edits, number of submissions made without compilation error, whilst submission behaviour considers if a student has tested or ran their code prior to submission.

In Koprinska et al. [15] decision trees were used to predict if students will pass or fail their final exam up to an accuracy of 87%. Data was collected from PASTA, a code submission website and Piazza which is an online question forum. The attribute practical_quiz mark was found to have the most information gain and was the root of the decision tree. Additionally, starting and finishing assessments early and reading the posts on Piazza contributed to the classification of a high predicted course mark, both in the mid semester and end of semester clustering.

The score and performance of students of different learning styles has been explored by McBroom et al. [16]. Data was used from a university coding course including their coding homework submissions over the 13 weeks which were graded using testcases. By identifying student submission behaviour and linking submission behaviour to a student's performance, they were able to cluster students into six behavioural clusters using multidimensional vectors as input into the K-means algorithm. Features used in initial clustering included percentage of early, normal and late attempts, number of tests passed on first and last attempt, number of attempts and time taken between first and last submission. These clusters were then compared to five different grade classifications of F for fail (marks less than 50) P for pass (mark between 50 and 64), CR for credit (marks between 65 and 74), D for distinction (marks between 75 and 84) and HD for high distinction (marks above 85).

Demographic features have also been explored by McBroom et al. [17] where they explored gender and age in student performance of programming courses. The study utilised the Grok learning platform, specifically the 2018 National Computer Science School (NCSS) course for five python programming courses of students from Year 5 to Year 11. In most year groups, males and females completed the same number of exercises, however in Year 10 females completed more beginner exercises and males completed more intermediate exercise. Year 10 males had the highest task completion of all intermediate male groups whilst female students had slightly higher levels of difficulty in the beginning of the intermediate courses. Overall, only minor differences were found between male and female students in the study suggesting gender is not a leading influence in student performance in coding subjects.

Attributes can also be selected from student's submissions rather than based on student behaviour. In the case of Macedo et al. [18], their research clustered students in a Portuguese language course based on student's grammatical errors. The types of language errors were used to cluster students and it was observed that grammatical error relating to Nominal Agreement grammatical errors and

Verbal Agreement grammatical errors occur simultaneously. This was confirmed by a language expert who acknowledged teachers observe both of these errors at the same time in students during language classes.

2.6.1 Summary

From Koprinska et al. [15] it is reasonable to assume that students who engage in the learning content through asking questions or viewing the discussion boards show a high level of engagement in the material which influences the assessment marks. This finding can be applied to the dataset used for this thesis because the online forum posts in Koprinska et al. [15] are similar to the learning slides on Grok Learning which introduce the coding concepts before the coding problems. Since these learning slides are optional in Grok Learning it reflects the optional Piazza content in Koprinska et al. [15] and therefore viewing and interaction with the Grok Learning slides should be considered as a potentially notable feature during clustering of the thesis dataset.

McBroom et al. [16] showed general trends of students in clusters with high pass rates. Students with high pass rates usually start tasks early or in moderate time and students in clusters of low pass rates started late or not at all. This suggests that there are groups of high and low achieving students which can be identified early, based on their submission time or start time when working on courses. Additionally, McBroom et al. [16] showed that whilst behavioural characteristics of students are flexible, student learning style will generally be consistent overall.

In summary, past research has utilised a range of attributes for student behavioural clustering, and these attributes can primarily be grouped into time-based features, submission features and features relating to the accuracy of a submission.

2.7 Methodologies

2.7.1 Methodologies for student behaviour

Existing student behaviour clustering models include McBroom et al. [19] and their novel clustering algorithm called DETECT which allows for analysis of student behaviour in coding courses over time. DETECT stands for Detection of Educational Trends Elicited by Clustering Time-series data and is a hierarchical clustering technique which produces a clustering structure similar to a decision tree. A primary finding was the number of autosaves heavily influenced student performance as well as the time from first submission fail to completion. Autosaves are automatically triggered every 10 seconds, therefore a high number of autosaves and time from first fail to completion indicates a large number of pauses, likely due to a student who is struggling with the problem.

Macedo et al. [18] clustered students based on grammatical errors in two different semesters of an undergraduate Portuguese language course to find similarities in the attributes of the students with similar grammatical errors. Both K-means and Fuzzy-C means (FCM) algorithms were used for clustering. K-means utilised the distance between a datapoint and a cluster centroid to determine cluster belonging whilst Fuzzy clustering allows data points to belong to more than one cluster. The quality of the clusters was measured using intra-cluster dissimilarity and inter-cluster dissimilarity with Euclidean distance. The intra-cluster dissimilarity was calculated using the mean of the sum of distances between the data and the centroids whilst the inter-cluster calculation consisted of the average of the distances between each pair of centroids. For K-means, increasing the number of clusters had little to no impact on the inter or intra dissimilarity between clusters whilst FCM showed significant improvement with each increase in cluster from 2-9 clusters.

2.7.1.1 Summary

The DETECT algorithm [19] approach would be well suited to temporal data over a time series and allows for easily interpretable results. However, the disadvantages of DETECT is it is similar to hierarchical clustering and the algorithm may not perform well on high dimensionality data.

Additionally, [18] highlights the importance of clustering 'goodness' measures to check for separation and cohesion of the produced clusters. The FCM algorithm shows an advantage over K-means as it can automatically determine the number of clusters, can process overlapping data and is robust to initialisation, noise and outliers. In comparison, the disadvantage of K-means is its high dependence on initialisation as well as the inability to cluster overlapping data well. To help improve K-means initialisation of the *k* number of clusters, methods such as the Elbow method and Silhouette Method can be used to adjust K-means' initialisation and allow for insight into the number of clusters needed, especially over larger datasets.

2.7.2 Methodologies for student submissions

Clustering of student submission mistakes includes using programming submission syntax to cluster similar submissions. Perry et al. [20] created a clustering algorithm called CLARA for online coding submissions which clusters python programs based on syntax. CLARA produced 51 clusters for a programming assignment of 200 submissions and 4 solution strategies. The algorithm aimed to use correct student submissions to repair incorrect student submissions and used clustering in two ways. The first was to cluster correct submissions and remove 'dynamically equivalent' submissions which would produce the same outputs. Clustering was also used on incorrect student attempts to run a repair algorithm against all clusters of incorrect student attempts and find a minimal repair for the candidates. Whilst the algorithm was able to repair up to 97% of incorrect submissions, the program was run on small to medium sized programming submissions. Additionally, the algorithm was successful at handling some complex programming submissions including multiple functions, multiple loops and nested loops.

Other studies utilised domain experts to confirm the clustering of students based on their programming submission. Kaleeswaran et al. [21] used a semi-supervised method of clustering submissions. Submissions were clustered by solution strategy and instructors could then identify or add correct submissions into each cluster. Submissions are then checked against the instructor validated submissions in the same cluster. This method was applied to dynamic programming assignments to provide verified feedback for 85% of submissions. Whilst expert knowledge improved the accuracy of the clusters, the reliance on expert knowledge is a limitation of the algorithm as the algorithm requires manual human input which may not be possible in cases of large numbers of clusters.

Recommendation and feedback algorithms have also been developed including Sharma's et al. [22] novel TipsC algorithm. TipsC is a recommendation algorithm for C language coding submissions that searches for correct code submissions similar to the incorrect submission to suggest recommendations to improve the current code. It can help resolve logical runtime errors and also act as a cluster visualisation tool for teachers and coordinators. TipC uses Abstract Syntax Trees to normalise C programs into a linear representation based on a modified version of Levenshtein's distance and then similar correct programs are clustered using hierarchical clustering.

Programming submissions can also be analysed using network graphs as in McBroom et al. [23] where they developed Progress Networks, inspired by intersection networks to identify student mistakes in coding submissions. Six different Grok Learning NCSS problems were used as data inputs. The code submissions were transformed into a trace which is a series of successive states and then generated into a network. The states are represented as nodes in the Progress Network whilst the transitions are edges. To ensure state matching was, abstract syntax trees were used to remove comments and normalise variable names. This allowed for identification and grouping of similar student program submissions which are written differently but perform the same task. The order of the states was created in terms of the testcases completed and gives insight into a student's progress on a programing problem.

The Progress Networks were compared to each other through normalising the edge counts by the total number of submissions and matching nodes between networks. Differences in normalised edge count was also created to show the frequency of which edges were used most by students thereby showing the number of state changes of how a student completed their programming submission. The comparison of the networks allowed for comparison of different grade levels and different programming questions, and it was found that students performed better on similar problems. Results also showed a lack of state improvement was more common for younger students who generally stayed on the same failed testcase after a following submission.

2.7.2.1 Summary

Of the algorithms explored, Progress Networks [23] would provide meaningful insight into the progress made by students in different types of questions and would also be useful to compare different clusters of students and their progress on programming submissions compared to their peers.

2.7.3 Methodology for correlation of variables

Existing research has considered how different behavioural features correlate to learning performance. Cerezo et. al. [24] sought to determine how procrastination behaviours and time spent on learning content was correlated to final exam marks. Using an ANOVA test with final marks as the dependent variable and clusters as the independent variable, statistical significance was shown between the four clusters. T-tests were also used to determine if the differences between the clusters were statistically significant.

To determine the significance of the clusters Li and Tsai [7] used Kruskal-Wallis tests to further identify student behaviour regarding the number of tasks completed, average time spend on each tasks and total time spent on all tasks of their three clusters. It was shown that students had different learning behaviours when viewing online learning materials based on the time spent on each view and viewing of announcements.

2.7.3.1 Summary

In both studies, statistical methods were used to identify differences in student behaviour within clusters and was able to prove correlation of features. Therefore, the ANOVA and Kruskal-Wallis test allowed for greater insight into the learning style of students within clusters and can be used alongside other measures of cluster similarity and separation to determine cluster quality.

Chapter 3 Methodology

3.1 Data

All data sourced for this thesis is from the Python beginner stream of the 2019 Grok NCSS programming course. This course was selected due to the high number of students and the diverse pool of students enrolled in the beginner stream, ranging from first time programmers to accomplished programming students.

The course consists of 5 weeks of content and graded problems. As described in Section 1.2.3, the course content slides, introduces students to the Python concepts which are tested in the graded problems. The slides also can contain programming questions for students to practice their programming skills, where they are allowed unlimited attempts without being penalised. There are 8 scored problems for each of the 5 weeks of the course and these problems are marked by auto-grading testcases. The highest possible score for the course is 400 with a maximum of 10 possible marks awarded for each correct question. The collated data includes both problem submission data as well as navigational data on how students utilise the Grok platform.

The beginner 2019 Python course has 10,558 students. Vectors were created for each student for each of the 5 weeks of the course, resulting in a total of 52,790 vectors. One vector was created for each week regardless of if a student had any course activity for that week. Any late submissions or course activity for a student after the deadline was ignored.

3.2 Algorithms

3.2.1 K-means algorithm

The data has been processed to be used as an input into K-means clustering algorithm. Clustering was used due to clustering being an unsupervised method [2] which is appropriate for the unlabelled Grok NCSS data in this thesis. Unlabelled data refers to data that does not have a classification label. An example of classified data in an online programming learning context is students who are labelled as low achievers, medium achievers and high achievers based on their scores. Since K-means will be used to group students into similar behavioural or procrastination clusters, currently there is not a behavioural or procrastination classification that exists for each student within the dataset. Therefore, the K-means algorithm will produce clusters of students that can then be analysed and classified based on their similarities and centroids.

K-means is a clustering algorithm which has *k* as an input, indicating the number of clusters to produce [25]. The algorithm then selects *k* random points, and these points can either be from the original dataset or produced based on the input data without being a data point in the original dataset [25]. The python K-means functionality uses random points which are not necessarily in the input dataset [26]. These *k* points become the centroids of the clusters and each data point is assigned to the closest cluster centroid based on a distance function. After all data points are assigned to a cluster the centroids are then recomputed. The K-means algorithm in this research completes 10 iterations of K-means with different centroid seeds prior to outputting the final result [26].

3.2.2 Elbow method

The Elbow method will be used in conjunction to K-means clustering to determine the number of clusters to generate from the K-means algorithm [25]. The Elbow method runs the K-means algorithm for a number of *k* values. For each value of *k* the distortion score is calculated which is the sum of square distances from each point to the centroid of the cluster [27]. The distortion score is then visualised as a line graph and the 'elbow' inflection of the curve indicates the optimal value of *k* for the dataset [27]. The Elbow method was used to determine the number of clusters to input into the K-means algorithm for the final results presented in this thesis.

3.2.3 Markov chains

Following the clustering of data using K-means, Markov chains will be used to determine student's learning style changes and how students move between clusters each week. Markov chains are a stochastic model and show the probability for consecutive series of events [28]. The events are referred to as states and the probability of the transitions between each state is visualised as a directed graph, with the states as nodes, and the probabilities as edges between each node. The probabilities are solely based on the present state, and do not consider the history prior to arriving to that state [28]. To represent student learning changes, Markov chains will be used to show the percentage of students who began in a cluster in Week 1, and then transitioned to another cluster in the following weeks of the course. Markov chains will also be used for each consecutive week from Week 1 to Week 2, from Week 2 to Week 3, from Week 3 to Week 4 and from Week 4 to Week 5 with a total of four Markov chain consecutive week transitions for each set of clusterings.

3.2.4 Information gain

Information gain (IG) is used in this thesis to indicate cluster membership by ranking the features based on their information gain value. Information gain calculates the reduction in entropy by splitting a dataset on a randomly selected feature and subtracting the entropy of this split from the starting entropy on the dataset [29], where entropy is the amount of variance in the data. Features with a high information gain indicate the cluster is highly characterised by the feature. For this thesis, information gain is calculated on each feature using binary cluster flags where the features of one cluster is compared to all other clusters, for example Cluster 0 features are compared against features in Cluster123 representing Cluster 1, 2 and 3.

3.3 Python implementation of algorithms

Python was used to facilitate data pre-processing as well as implementing the algorithms used. The data vectors were created by converting the json files with the raw data into Pandas DataFrames in Python. DataFrames allowed for querying and manipulation of data when calculating the features for each feature vectors [30]. A module was created with Python code which consisted of functions for data pre-processing and the programming logic for calculating the features in each vector. For K-means clustering, the scikit Elbow method was used to output the optimal number of clusters, and the scikit K-means function was used to generate the centroids of each cluster. Markov chain calculations were implemented using Python and then visualised using a network graphing module. Python was also used to generate the bar charts and pie chart visualisations for the scores and Markov chains.

3.4 Feature selection

Feature selection in this study draws from existing literature which utilises FSLSM behavioural features and procrastination features and applies them to an online programming course context. Each student was represented with a total of 5 vectors, one for each of the 5 weeks of the course.

3.4.1 FSLSM behavioural features

The FSLSM behavioural learning style features are inspired from Graf's et al. [10] research where Felder-Silverman's learning model [9] was used to investigate student's navigational behaviour in an online programming course. In the study, the students completed a questionnaire to self-report their FSLSM behaviour type, and this FSLSM category was used to compare the navigational approaches between students of the same behavioural category. The FSLSM theory describes eight different learning styles (active/reflective, sensing/intuitive, visual/verbal and sequential/global learners) and how each learning style indicates a student's navigational behaviour and interaction with the course website. One feature which was most indicative of each type of learner was selected for this thesis and adjusted to fit the Grok NCSS beginner course. All FSLSM learning groups will be used except for the visual and verbal learning styles due to the difficulty in measuring visual or verbal preference in an online course without student self-assessment, as student self-reporting of behaviour types will not be used in this study. It should be noted that while the FSLSM features have been informed by the FSLSM learning theory, it is not expected that the clusters will strictly conform to the FSLSM behaviour classifications.

Based on Graf et al. [10] findings, a number of hypotheses have been made on how each behavioural group's defining trait would be reflected in the Grok NCSS course in Table 3.1. These hypotheses have informed the FSLSM features selected for clustering as outlined in Table 3.2.

TABLE 3.1 FSLSM LEARNING BEHAVIOUR CATEGORIES AND HYPOTHESIS ON HOW THE FSLSM CATEGORIES WOULD BE EXPRESSED IN THE GROK NCSS COURSE

FSLSM behaviour learning category	Category description	Grok NCSS course behaviour hypothesis
Active learners	Learn by trying things in a hands-on way. Prefers working collaboratively.	Frequent programming submissions.
Reflective learners	Cerebral learners, learns through thinking and reflecting.	Takes longer on problems to reflect on feedback.
Sensing learners	Prefers concrete examples, practical learners, high attention to detail.	Completes majority of slide questions.
Intuitive learners	Prefers abstract concepts and are innovative. Attempts problems without completing previous questions.	Works on several problems at once.
Sequential learners	Learns in linear steps and follows course content sequentially.	High number of sequential slide views.
Global learners	Learns in large leaps and prefers a high degree of freedom in the learning process. Prefers learning from conclusions and summaries.	High number of course menu views to provide a high-level overview of the course.

TABLE 3.2 FSLSM BEHAVIOUR FEATURES

FSLSM learning style	Feature	Description
Active	Number of terminal runs	Number of times a student runs their problem program attempt.
Reflective	Average time spent on problem	Average time spent on problems for the week from first to last submission.
Sensing	Percentage of completed slides	Percentage of content slides with all questions answered correctly.
Intuitive	Problems attempted simultaneously	Represented as a binary with 0 for no problems attempted simultaneously and 1 for problems attempted simultaneously.
Sequential	Percentage of sequential slide views	Percentage of slide views which are immediately before or after the current slide viewed, out of all of the slides viewed by a student for the week.
Global	Number of course menu views	Number of times a student viewed the course menu overview.

3.4.2 Procrastination features

Other literature utilises procrastination features for clustering. Cerezo's et al. [11] research incorporates both time related feature and effort related features. Time procrastination features measured how long a student works on a component of the course or how long a student delays starting course questions, whilst effort procrastination features measure how much of the course a student completes and to what standard. Additionally, research by Hooshyar el at. [12] defines a procrastination time feature of Inactive Time, which refers to the time from when an assignment opens to when a student first views the assignment. This feature has been adapted to this study to refer to the time between when an assignment opens to when a student first submits an attempt to that problem. The selected procrastination features are outlined in Table 3.3 and include both time and effort related features.

Hooshyar et al. [12] also define another feature called Spare Time which is the time from when a student submits an assignment until the assignment due date. Whilst this would be an interesting feature inclusion, the nature of the Grok course deadlines would skew the spare time data for each week. Grok releases each week’s worth of content separately, and all of the problems are due in the final week, resulting in Week 1 spare time measures being significantly higher than spare time calculated in Week 5, since students generally submit in the week that the content is released. To continue this example, problems in Week 1 would be submitted in Week 1 but due in Week 5, whilst Week 5 content would be submitted and due in Week 5. This would skew the data as all spare time calculated from Week 1 would suggest very low levels of procrastination compared to spare time calculated in Week 5. Therefore, the Spare Time feature will be omitted from this thesis.

TABLE 3.3 PROCRASTINATION FEATURES

Feature	Description
Inactive time	Amount of time between when an assignment is released to a student’s first problem submission. Calculated in seconds.
Time spent on content	Amount of time from first slide view to last slide view. Calculated in seconds.
Average time spent on problems	Average time between first and last problem run for each problem in a week. Calculated in seconds.
Percentage of completed problems	Percentage of problems which passed all testcases.
Number of autosaves	Number of automatic autosaves. Autosaves are generated after periods of inactivity on the website whilst students are completing programming problems.
Number of terminal runs	Number of times a student runs their programming problem.
Percentage of slides completed	Percentage of slide problems completed. Slide problems are not scored and are questions relating to the concepts taught each week.
Percentage of slides viewed	Percentage of slides viewed by the student for the week.

3.5 Data pre-processing

3.5.1 Cluster data pre-processing

Prior to clustering the data to produce the results in Chapter 4, the following data pre-processing actions were undertaken: normalisation, removal of outliers and dealing with missing values. The data for FSLSM behaviour and procrastination features were normalised between a range of 0 to 1 for each feature. This was to reduce bias of features during clustering since K-means relies on a distance measurement which can be impacted by features measured on different ranges.

Outliers were also removed for some features for two reasons, the first is to improve the distinctiveness of the clusters as reflected by the Silhouette and Davies Bouldin score. The second reason was to increase the meaningfulness of the centroids. Some features calculated had a high range of values, but majority of values were in the lower end of this range, resulting in these features having a very low mean after normalisation. Features measured in seconds in particular had a long tail of high values with a range of 0-691,200 seconds, where the upper value represents time durations as large as 7 days. Majority of the time values were within the 2 day duration and therefore the outliers on the upper range of time features were removed to improve the interpretability of the mean and centroids of these features.

Missing values were handled during data pre-processing prior to clustering. Vectors with missing values were removed in both the behaviour and procrastination dataset due to the high number of missing values which would skew clustering if replaced with a filler value such as 0 or the mean for that feature. After removing missing values there were 15,733 FSLSM behaviour vectors and 19,724 procrastination vectors used for the final clustering.

3.5.2 Revised procrastination feature selection

A preliminary version of the clusters was implemented using K-means on the behavioural and procrastination features as listed in Table 3.2 and Table 3.3. However, it was noted the procrastination features had a non-optimal Silhouette and Davies Bouldin score. Using all procrastination features in Table 3.3 produced a low Silhouette Score of 0.326 and high Davies Bouldin score of 1.039 where a Silhouette Score of 1 is ideal and indicates little overlap of the clusters, and low Davies Bouldin score indicates unique clusters. To improve the cluster results, the procrastination features were split into two, the first set of features were time-based features relating to how long a student delayed starting course problems and how long it took for students to complete aspects of the course. The second set of features are effort related and measure how much a student engaged with the course. The updated features are outlined in Table 3.4 and Table 3.5. These two sets of features were clustered separately in the results described in Chapter 4.

TABLE 3.4 FINAL PROCRASTINATION TIME FEATURES

Feature	Description
Inactive time	Amount of time {seconds} between when an assignment is released to a student's first problem submission.
Time spent on content	Amount of time {seconds} from first slide view to last slide view.
Average time spent on problems	Average time {seconds} between first and last problem run for each problem in a week.
Percentage of completed problems	Percentage of problems which passed all testcases.

TABLE 3.5 FINAL PROCRASTINATION EFFORT FEATURES

Feature	Description
Percentage of completed problems	Percentage of problems which passed all testcases.
Number of autosaves	Number of automatic autosaves. Autosaves are generated after periods of inactivity on the website whilst students are completing programming problems
Number of terminal runs	Number of times a student runs their programming problem.
Percentage of slides completed	Percentage of slide problems completed. Slide problems are not scored and are questions relating to the concepts taught each week.
Percentage of slides viewed	Percentage of slides viewed by the student for the week.

3.5.3 Markov chain data processing

The Markov chains are concerned with showing student's transition between each cluster and student's drop-off rates. To represent drop-off rates, two new clusters were added, these clusters are No Attempt which means a student had no engagement with the course for that week, and Partial Attempt which indicates a student partially engaged with the course content. The No Attempt students had all empty vectors for the week and Partial Attempt students had a partially empty vector for the week. Previously empty vectors which were removed for clustering were added to the Markov chain calculations resulting in a total of 44,160 vectors for both the FSLSM behaviour and procrastination Markov chains. This is slightly lower than the original 52,790 vectors, which is because the students who were inactive and did not engage in any aspect of the course for any week were removed from the dataset.

3.6 Limitations

There are some limitations in the proposed methodology. A data related limitation is this proposed method only considers a student's course interaction per week and does not consider the overlap between the weeks. This only impacts two features in the FSLSM behaviour clusters, problems attempted simultaneously and percentage of sequential slide views. For example, sequential slide views only consider the slide views for a week's worth of slide content. However, a student could have viewed a slide from Week 2, then a slide from Week 1 before returning to the next slide in Week 2. Whilst this would have impacted the sequential slides viewed, we did not include inter-week interactions in the scope of the feature calculations.

An additional data limitation is the use of averages in the calculation of features. The vectors are calculated from a week's worth of course interaction, and therefore averages were required for features related to problems since there are 8 problems in each week. These averages could mask the wide range of ways students interacted with individual problems in each week. However, using weekly vectors was a necessary decision to prevent the dataset from becoming too large and introducing noise to the clusters. Therefore, whilst averages can limit the insight we can obtain from some features, this was a compromise made to reduce noisy data and increase the meaningfulness of the clusters overall.

K-means clustering has a limitation because the algorithm assumes that the variance of the distribution of each feature is spherical [2]. As a result, if the dataset does not have a spherical distribution, it will break an assumption of K-means and possibly result in sub-optimal cluster. Since the K-means algorithm is used for clustering in this thesis, there is an opportunity to explore other clustering techniques which do not rely in a spherical feature distribution in future research.

Markov chains have a limitation because they only consider the probability of an event occurring from the current state, and do not consider the state's history in its calculations. Whilst this allows for easier interpretability of the model due to its simplicity, it is possible the Markov chain may not adequately explain the complexities in the changes in student behaviour.

The feature selection and measurements are somewhat limited as they cannot adequately capture the external factors which may impact a student's learning style or performance. For example, a procrastination feature which measures a high time spent on problems for a student will not indicate other external factors such as other student commitments outside of the course or students multitasking with other non-NCSS course activities. Therefore, the insights from the study will speculate on a student's behaviour or procrastination tendencies, and the clusters will not be used as a definite indicator of student learning type.

3.7 Assumptions

The following assumptions were made during the processing of data and the proposed methodology:

- The methodology assumes that a student's behaviour and procrastination tendencies remain consistent for the week, and any changes in learning style only occur between each week's worth of content. In reality, a student's learning style may change regularly between each question and this proposed methodology will only measure a weekly level of granularity in the changes in student learning style.
- Time related procrastination measures assume a student with a high delay time or high duration of time for completing a task is procrastinating. However, a student may have other commitments preventing them from starting tasks early or completing them quickly.
- We assume that whilst there will be students of differing FSLSM behaviour types in the Grok NCSS data, we cannot assume that all FSLSM categories will be represented in this

dataset. Since this thesis will not include a student survey for self-reporting of their behavioural type, the distribution of students who would fit into each FSLSM behavioural classification is not clear. Therefore, the behavioural clusters cannot be directly compared to the FSLSM theory. Thus, the clusters generated by the FSLSM behaviour features will be interpreted in light of the FSLSM literature, but likely will not exactly reflect the 6 behavioural groups measured.

- Students who work on problems after a submission was passed successfully are doing so optionally to further their learning. Therefore, we have not included any student data logged after a problem was successfully completed with all testcases passed. Similarly, any student data logged after the course deadline was also omitted.

Chapter 4 Results

4.1 Objective 1 results: Investigate similarities in student learning style in terms of learning behaviours and procrastination tendencies

Similarities in student behaviour were analysed in this thesis in two ways: using the FSLSM behavioural features and using the procrastination features. Three separate K-means clusterings were used, one on the FSLSM behavioural features and two on the procrastination features which are split into time features and effort features.

4.1.1 FSLSM behaviour clustering

FSLSM behaviour clustering results produced 4 vectors, as informed by the Elbow method for this data. The distribution of vectors between each cluster is shown in Table 4.1, the centroids of the clusters are shown in Table 4.2 and the information gain of the features in each cluster is in Table 4.3. Outliers were removed from features to improve the accuracy of results and to increase the meaningfulness of the centroids. Despite outliers being removed, for time features like average time spent on problems, there is still a very large range of values from 0 seconds to 518400 seconds which is equivalent to 6 days. The majority of the time values are on the lower end of this range, which has resulted in the centroids for time values being small after normalisation.

TABLE 4.1 DISTRIBUTION OF VECTORS IN FSLSM BEHAVIOUR CLUSTERS

	Number of vectors	Percentage of vectors
Cluster 0	3485	22.15%
Cluster 1	2759	17.54%
Cluster 2	8598	54.64%
Cluster 3	891	5.67%
Total vectors	15733	

TABLE 4.2 FSLSM BEHAVIOUR CLUSTER CENTROIDS

		Cluster			
	Full data	0	1	2	3
Terminal runs	0.122	0.087	0.199	0.108	0.154
Slides completed	0.671	0.246	0.849	0.831	0.241
Consecutive slides viewed	0.745	0.726	0.727	0.765	0.679
Menu views	0.114	0.074	0.196	0.101	0.134
Average time spent on problems	0.087	0.083	0.182	0.049	0.171
Simultaneous problems	0	0	1	0	1

TABLE 4.3 FSLSM BEHAVIOUR FEATURES RANKED BASED ON INFORMATION GAIN

Cluster 0		**Cluster 1**		**Cluster 2**		**Cluster 3**	
Feature	*Gain ratio*	*Feature*	*Gain ratio*	*Feature*	*Gain Ratio*	*Feature*	*Gain ratio*
Simulatneous problems	0.474	Simulatneous problems	0.619	Simulatneous problems	0.420	Simulatneous problems	0.164
Slides completed	0.125	Slides completed	0.042	Slides completed	0.151	Slide completed	0.100
Average time spent on problems	0.022	Average time spent on problems	0.042	Average time spent on problems	0.027	Average time spent on problems	0.012
Consecutive slide views	0.016	Terminal runs	0.030	Consecutive slide views	0.023	Consecutive slide views	0.008
Terminal runs	0.012	Course menu views	0.025	Terminal runs	0.016	Terminal runs	0.005
Course menu views	0.011	Consecutive slide views	0.011	Course menu views	0.009	Course menu views	0.002

From the cluster centroids in Table 4.2 and the information gain scores in Table 4.3, we can gain the following insights.

Cluster 0: This cluster is characterized by the lowest number of terminal runs (Table 4.2) which represents the number of runs a student made of their program solution before submitting their solution for auto-graded testcase marking. Despite viewing 72.6% of the slides consecutively on average, students in this cluster only completed 24.6% of slides on average. The high number of consecutive slide viewership would be consistent with a Sequential learning style. However, the low number of slides completed may indicate that students primarily read the content slides instead of interacting with the slide questions, or students may have skimmed through the slides without reading them. The information gain for slides completed is 0.125 for Cluster 0 (Table 4.3), which is the second highest information gain for this feature compared to all clusters. Therefore, it can be concluded Cluster 0 is characterised by a low slide completion rate.

Cluster 1: This cluster has the highest number of completed slides with an average of 84.9% of slides completed successfully. However, Table 4.3 shows that this cluster's membership is not predicted by a high percentage of slides completed as it has the lowest information gain for this feature compared to the other cluster's information gain for the slides completed feature. Cluster 1 also has the highest number of course menu views which is a hallmark indicator of Global learners and the highest number of terminal runs which is an indicator of Active learners. The students in this cluster also worked on more than one problem at once which can indicate Intuitive learning style. This is further supported by Table 4.3 which shows that Cluster 1 had the highest information gain on the number of problems worked on simultaneously. These students also work the longest on problems on average, which may be due to these students working on more than one problem at once.

Cluster 2: Similarly, to Cluster 1, this cluster also has a high number of completed slides with 83.1% of slides completed. However, Cluster 2 had high information gain compared to the other clusters for this feature, which suggests Cluster 2 membership is characterised by high slide completion rates which describes the Sensing learning style. Whilst this cluster is similar to Cluster 1 regarding their centroids, the main difference is evident in Cluster 2 having the lowest amount of time spent on problems, and that on average students complete one problem at a time. Cluster 2 also has the highest number of vectors and consisted of 54.64% of the overall data.

Cluster 3: Cluster 3 has a high number of terminal runs and time spent on problems, indicating these students have a high engagement with the problems. However, Cluster 3 is not engaged with the content information and has the lowest slide completion and slide views. They also have high course menu views which could indicate skipping to problems and not engaging with the slide content which would be suggestive of a Global learning style. Whilst Cluster 3 has a high average time spent on problems, the low information gain for this feature relative to other clusters indicates that average problem time does not necessarily indicate Cluster 3 membership.

4.1.2 Time procrastination clusters

Time features outlined in Table 3.4 produced 4 clusters for K-means clustering, where the *k* value was based on the output from the Elbow method. Table 4.4 shows the distribution of the vectors in each cluster, Table 4.5 shows the cluster centroids and Table 4.6 indicates the information gain of each feature in the clusters.

TABLE 4.4 DISTRIBUTION OF VECTORS IN TIME PROCRASTINATION CLUSTERS

	Number of vectors	Percentage of vectors
Cluster 0	3170	16.07%
Cluster 1	8182	41.48%
Cluster 2	2749	13.94%
Cluster 3	5623	28.51%
Total	19724	

TABLE 4.5 TIME PROCRASTINATION CLUSTER CENTROIDS

		Cluster			
	Full data	0	1	2	3
Inactive time	0.043	0.077	0.044	0.024	0.031
Time spent on content	0.264	0.166	0.088	0.701	0.357
Percentage of completed problems	0.876	0.349	0.980	0.968	0.976
Average time spent on problems	0.054	0.012	0.003	0.006	0.005

TABLE 4.6 TIME PROCRASTINATION FEATURES RANKED BASED ON INFORMATION GAIN

Cluster 0		Cluster 1		Cluster 2		Cluster 3	
Feature	*Gain ratio*	*Feature*	*Gain ratio*	*Feature*	*Gain Ratio*	*Feature*	*Gain ratio*
Completed problems	0.623	Time spent on content	0.341	Time spent on content	0.605	Time spent on content	0.374
Inactive time	0.023	Completed problems	0.183	Completed problems	0.025	Completed problems	0.093
Time spent on content	0.020	Average time spent on problems	0.032	Inactive time	0.016	Average time spent on problems	0.015
Average time spent on problems	0.018	Inactive time	0.016	Average time spent on problems	0.011	Inactive time	0.009

From the results we can obtain the following insights on the clusters.

Cluster 0 - High procrastination: Cluster 0 students had high inactive time which meant students waited longer to start problems after they were released compared to the other clusters. They also had the lowest number of completed submissions with only 34.9% of submissions completed. Table 4.6 shows Cluster 0 has the highest information gain on the percentage of completed problems feature which indicates that a low percentage of completed problems will likely result in Cluster 0 membership. The high amount of time spent on problems can indicate that these students struggled with the content which is why they had a low submission completion rate. Therefore, we can consider this group to have high procrastination tendencies.

Cluster 1 - Low procrastination: This cluster shows little evidence of procrastination because it has the lowest amount of time spent on content and problems whilst achieving a 98.0% problem completion rate. Time spent on content and problem completion rate were the highest ranked features for Cluster 1 indicating that Cluster 1 membership is characterised by a low amount of time spent on content and a high percentage of completed submissions. This could indicate this group had prior programming experience, allowing them to complete submissions quickly without heavy reliance on content or problem feedback.

Cluster 2 - Low procrastination: Cluster 2 started problems shortly after the release time and had the highest amount of time spent on content. Table 4.6 indicates that Cluster 2 can be characterised by a high amount of time spent on content with this feature having the highest information gain of any feature of any cluster. This cluster also had a high amount of completed submissions with 96.8% of submissions completed. Whilst a high amount of time spent on content could indicate high levels of procrastination, the other features of low inactive time and high number of completed problems indicate this cluster has low levels of procrastination and students take the time to engage with content slides in the course.

Cluster 3- Moderate procrastination: This cluster has a high number of completed submissions and a medium amount of time spent on content. Cluster 3 also has a moderate level of inactive time compared to the other clusters. The centroids and information gain do not indicate either strong procrastination or strong non-procrastinating features. Therefore, this cluster can be said to be moderate procrastinators.

4.1.3 Effort procrastination clusters

Whilst procrastination features are typically time related, Cerezo's et al. [11] procrastination clusters also drew from other effort related features to measure the level of student course engagement that occurred during the time students were on the platform. The Elbow method indicated k=5 clusters to be used for this set of K-means clustering, with the distribution of the clusters shown in Table 4.7, centroids in Table 4.8 and information gain in Table 4.9.

Table 4.7 Distribution of vectors in effort procrastination clusters

	Number of vectors	Percentage of vectors
Cluster 0	12135	61.52%
Cluster 1	2596	13.16%
Cluster 2	1808	9.17%
Cluster 3	1443	7.32%
Cluster 4	1742	8.83%
Total	19724	

Table 4.8 Effort procrastination cluster centroids

		Cluster				
	Full data	0	1	2	3	4
Number autosaves	0.089	0.998	0.089	0.039	0.081	0.067
Terminal runs	0.057	0.064	0.058	0.025	0.050	0.048
Completed problems	0.876	0.983	0.973	0.276	0.461	0.947
Slides completed	0.647	0.839	0.267	0.292	0.727	0.176
Slides viewed	0.842	0.962	0.864	0.412	0.823	0.429

TABLE 4.9 EFFORT PROCRASTINATION FEATURES RANKED BASED ON INFORMATION GAIN

Cluster 0		Cluster 1		Cluster 2		Cluster 3		Cluster 4	
Feature	*Gain ratio*	*Feature*	*Gain ratio*	*Feature*	*Gain Ratio*	*Feature*	*Gain ratio*	*Feature*	*Gain ratio*
Completed problems	0.242	Completed problems	0.305	Slides viewed	0.175	Completed problems	0.167	Completed slides	0.162
Completed slides	0.231	Slides viewed	0.117	Completed slides	0.077	Completed slides	0.029	Slides viewed	0.041
Slides viewed	0.169	Completed slides	0.108	Completed problems	0.022	Slides viewed	0.026	Completed problems	0.033
Terminal runs	0.041	Terminal runs	0.048	Number autosaves	0.008	Number autosaves	0.003	Terminal runs	0.017
Number autosaves	0.017	Number autosaves	0.032	Terminal runs	0.006	Terminal runs	0.003	Number autosaves	0.016

The effort procrastination clusters are described below.

Cluster 0 - High problem effort, high content effort: Cluster 0 has the highest amount of engagement on all measures, with 98.3% of problems completed, 83.9% of slides completed and 96.2% of slides viewed. They also had a high number of terminal runs and autosaves which suggests they spent a significant amount of effort on the problem component of the course. Cluster 0 has a high information gain on both problem and slide features with completed problems and completed slides being the highest information gain for this cluster. Therefore, students from this cluster are highly engaged with both the problems and the content of the course.

Cluster 1 - High problem effort, medium content effort: This cluster is characterised by a low number of slides completed despite having a high number of slides viewed. Students viewed 86.4% of slides but only completed 26.7% of slides on average. This suggests students in this cluster may skim through the content without engaging with the practice questions. The information gain feature rankings for this cluster indicates that slide viewership is more characteristic of this cluster than slide completion rates. Since these students are somewhat engaged with viewing the content but do not answer the content questions whilst still achieving high problem completion rates, these students can be considered highly engaged with the problems but moderately engaged with the course content.

Cluster 2 - Low problem effort, low content effort: Cluster 2 has low activity on all features. The highest information gain for this cluster is for slides viewed which indicates Cluster 2 is characterised by low slide viewership of around 41.2%. These students likely didn't engage with the course content or problems and have the lowest number of completed problems.

Cluster 3 - Low problem effort, high content effort: This group of students had a low number of completed submissions, on average only completing 46.1% of the problems. The information gain for completed submissions is the highest for this cluster, suggesting that low problem completion is a strong characteristic of the students in this cluster. Despite the low problem engagement, students had high slide viewership and the highest slide completion rate with over 72.7% of slides completed.

Cluster 4 - High problem effort, low content effort: These students had a low number of autosaves and terminal runs but had a high completion rate of problems with an average of 94.7% of problems completed. Whilst a low number of autosaves and terminal runs could indicate a lack of engagement with the course problems, the high number of completed problems for this groups indicates that this cluster of students may have been quicker and more efficient at solving problems, perhaps because of prior programming knowledge obtained before completing the course. Based on the information gain feature rankings, this cluster can be characterized by a very low amount of slide content interactions, with only 17.9% of slide questions completed and 42.6% of slides viewed. Therefore, this cluster can be considered high problem effort, but low content effort students.

4.1.4 Cluster quality

The Silhouette and Davies Bouldin score can indicate the health of the clusters regarding their cohesion and similarity. An optimal clustering would have high cohesion meaning high similarity within clusters, and high separation meaning low similarity between different clusters [2]. The ideal Silhouette score is 1 which indicates little overlap and high separation of clusters, and Silhouette scores have a range of -1 to + 1 [31]. Conversely, a low Davies Bouldin score is optimal and indicates unique and highly cohesion clusters, where the lowest score is 0 [32]. Table 4.10 shows the Silhouette and Davies Bouldin scores for each of the three clusterings.

TABLE 4.10 SILHOUETTE SCORE AND DAVIES BOULDIN SCORE FOR FSLSM BEHAVIOUR CLUSTERS, PROCRASTINATION TIME CLUSTERS AND PROCRASTINATION EFFORT CLUSTERS

Clustering	Silhouette Score	Davies Bouldin Score
FSLSM behaviour	0.526	0.887
Procrastination time	0.456	0.697
Procrastination effort	0.554	0.869

The FSLSM behaviour cluster scores indicate there is a medium level of cluster separation from the Silhouette score, and the Davies Bouldin score implies there is medium to low cohesion within each cluster. These scores are similar to the procrastination effort cluster which also had medium level of separation and medium to low cluster cohesion. The procrastination time had a slightly lower Silhouette score which suggests lower cluster separation, but also had a lower Davies Bouldin score which indicates it has the highest amount of cluster cohesion of the three clusterings. Whilst there is no single best clustering for both high separation and high cohesion, these scores show that the behavioural clustering has similar cluster goodness compared to the procrastination clusters.

Additionally, whilst the scores may not indicate high cohesion or separation, these scores do not determine the usefulness of these clusters to identify different types of learning styles in students. The analysis of the centroids and information gain show that the clusters produced are meaningful and worthwhile clusters for educators to use for classifying students based on their behavioural or procrastination tendencies.

4.1.5 Cluster overlap

Whilst undertaking separate clusterings can highlight the different groups of students regarding their FSLSM behaviours or levels of procrastination, investigating the overlap between the clusters can provide a greater understanding of how students undertake the course holistically.

The behavioural clusters and procrastination time clusters were selected for clustering overlap analysis because they best fit the two theories from previous literature and because they have the same number of clusters (k=4) allowing for more seamless analysis.

TABLE 4.11 NUMBER OF VECTORS IN PROCRASTINATION TIME CLUSTERS WHICH OVERLAP WITH FSLSM BEHAVIOUR CLUSTER VECTORS

		FSLSM behaviour clusters			
		Cluster 0	Cluster 1	Cluster 2	Cluster 3
Procrastination time clusters	Cluster 0	1040	154	748	83
	Cluster 1	1371	1233	3721	473
	Cluster 2	364	418	1510	110
	Cluster 3	701	954	2622	225

Table 4.11 shows the largest overlap of the clusters is between FSLSM behaviour Cluster 2 and procrastinate time Cluster 1 (low procrastination), Cluster 2 (low procrastination) and Cluster 3 (moderate procrastination). This may be because FSLSM behaviour Cluster 2 consists of 55.67% of all the vectors for that clustering. Procrastination time Clusters 1, 2 and 3 have a high number of submissions completed but varying amount of time spent on problems and content. This can suggest that students of behaviour Cluster 2 are also low to mid-range procrastinators.

Procrastination time Cluster 0 (high procrastination) had the lowest scores, started problems late, had a high amount of time spent on problems and a low problem completion rate. This cluster is mostly overlapped with behaviour Cluster 0. Behaviour Cluster 0 had low number of terminal runs as well as low slide completion. This could indicate that whilst students in procrastination time Cluster 0 spent a long time on problems, most of this time was not spent on terminal runs and therefore students may not have been running their Python program prior to submission.

By considering the overlap of both sets of clusters, we can gain greater insight into the behaviours of students of different procrastination levels. This can help teachers identify not only which behaviours can lead to procrastination but also how to best remediate this procrastination behaviour.

4.2 Objective 2 results: Examine how learning styles and levels of procrastination influence student's overall course score

In this results section of the thesis, the previous clusters will be further analysed regarding the student's score distribution of each cluster. The maximum score for a student is 400 and each problem is worth 10 marks each if a problem is completed and passes all testcases. This section will be followed by a conclusion on which clustering features led to the most meaningful insight into student performance in the NCSS Grok course.

4.2.1 FSLSM behaviour scores

From the clusters, we can determine how student learning behaviour impacts a student's score. Figure 4.1 and Table 4.12 shows the score distribution across the clusters as a percentage.

FIGURE 4.1 PERCENTAGE OF FSLSM BEHAVIOUR CLUSTERS IN EACH SCORE RANGE

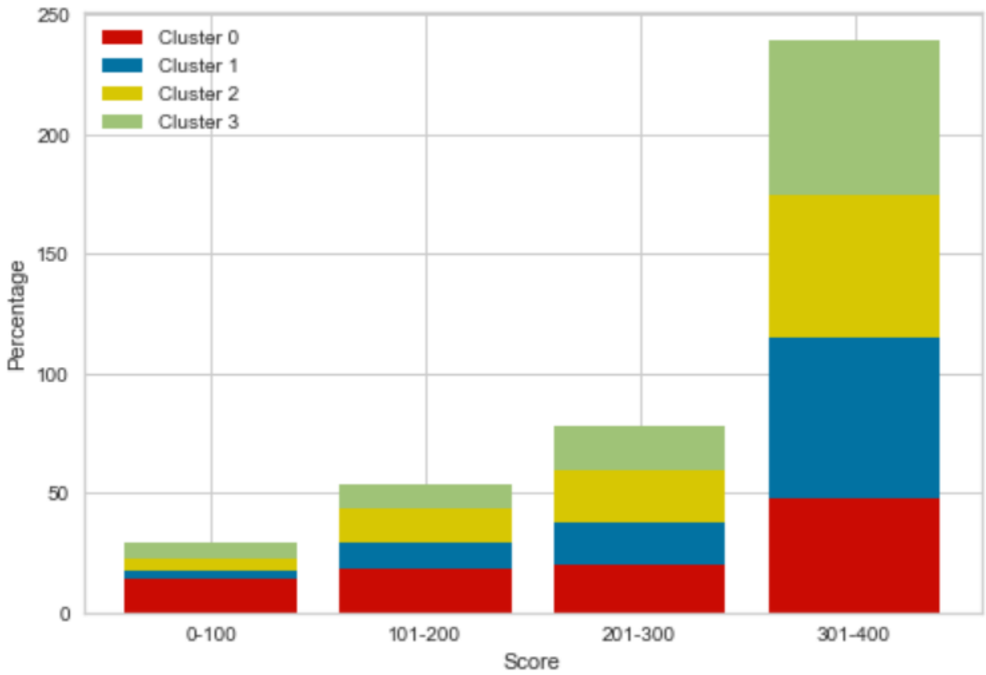


TABLE 4.12 PERCENTAGE OF FSLSM BEHAVIOUR VECTORS IN EACH SCORE RANGE

		Percentage of vectors in each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	3482	14.47%	18.04%	19.70%	47.79%
Cluster 1	2759	3.16%	11.45%	18.12%	67.27%
Cluster 2	8601	4.98%	13.99%	21.68%	59.35%
Cluster 3	891	6.62%	10.21%	18.41%	64.76%

It is evident that each cluster has a similar distribution of students of all performance types including low, medium and high performing students. Each cluster has a similar score distribution with the exception of Cluster 0 which has a significantly lower number of students in the highest score range of 301-400, with only 47.79% of student obtaining this mark compared to 59.35%-67.27% of students obtaining this mark in other clusters (Table 4.12). Therefore, the FSLSM features would not be a strong predictive measure of score, as also indicated by the medium to low correlation coefficient for the behavioural clusters of 0.349 in Table 4.15. However, this score distribution reflects the literature on the FSLSM learning theory which states that any FSLSM student groups aim to describe student learning behaviour and are not indicative of student performance.

4.2.2 Procrastination time scores

The procrastination time score distributions are outlined in Figure 4.2 and Table 4.13.

FIGURE 4.2 PERCENTAGE OF PROCRASTINATION TIME CLUSTERS IN EACH SCORE RANGE

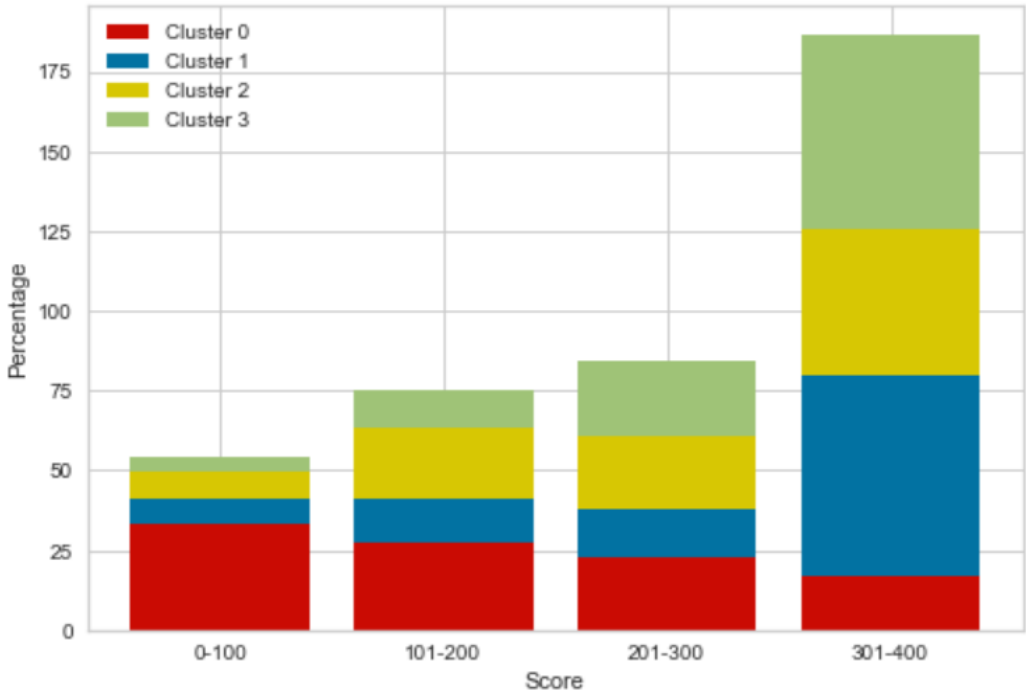


TABLE 4.13 PERCENTAGE OF PROCRASTINATION TIME VECTORS IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	3170	33.34%	27.35%	22.68%	16.63%
Cluster 1	8706	7.21%	13.81%	15.45%	65.53%
Cluster 2	2717	8.83%	22.34%	23.12%	45.71%
Cluster 3	5131	4.52%	11.67%	23.66%	60.15%

From the procrastination time scores in Figure 4.2 and Table 4.13, we can see that all clusters except for Cluster 0 show a similar distribution of marks in the low, medium and high range. This score distribution supports the cluster centroids since Cluster 0 (high procrastination) has the lowest problem completion rate, with only 16.63% of scores achieved in the 301-400 range. Interestingly, whilst Cluster 2 (low procrastination) and Cluster 3 (moderate procrastination) had similar centroids for most features, Cluster 2 had more students in the low to mid score range, with 54.29% of vectors between 0-300 marks, whilst Cluster 3 had 39.85% of students in this same score range. The predominant difference between these two cluster's centroids is students in Cluster 3 spent less time on content and slides. This can indicate that Cluster 3, like Cluster 1 had students with previous programming knowledge who didn't have to rely on the slide content as heavily as other students whilst still achieving high scores.

Additionally, Cluster 3 (moderate procrastination) had 39.85% of students in the 0-300 range which is similar to Cluster 1 (high procrastination) which had 36.47% of students in the low to mid-tier score range. Therefore, despite Cluster 3 having higher procrastination tendencies they were still able to achieve similar results to students with low procrastination tendencies.

4.2.3 Procrastination effort scores

Procrastination effort scores are outlined in Figure 4.3 and Table 4.14. The results show Cluster 2 (low problem effort, low content effort) and Cluster 3 (low problem effort, high content effort) have the lowest number of students obtaining a high mark in the 301-400 range. This is consistent with the centroids of these two clusters which shows low problem completion for both groups.

FIGURE 4.3 PERCENTAGE OF PROCRASTINATION EFFORT CLUSTERS IN EACH SCORE RANGE

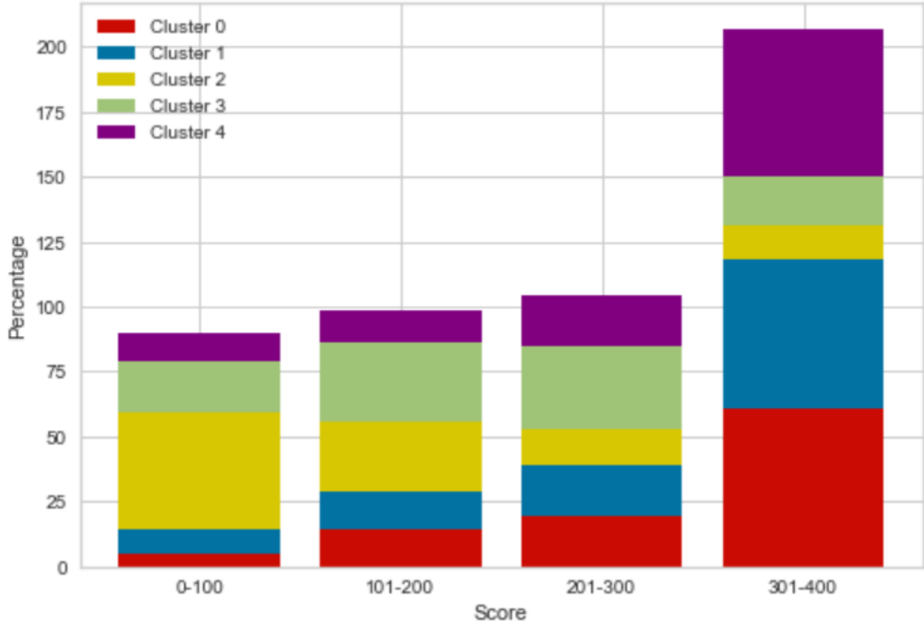


TABLE 4.14 PERCENTAGE OF PROCRASTINATION EFFORT VECTORS IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	12135	5.23%	14.46%	19.23%	61.08%
Cluster 1	2596	8.90%	14.21%	20.07%	56.82%
Cluster 2	1808	45.47%	27.43%	13.77%	13.33%
Cluster 3	1443	19.13%	29.94%	31.81%	19.12%
Cluster 4	1742	11.08%	12.80%	19.75%	56.37%

Three procrastination effort clusters have high achieving students in the highest score range: Cluster 0 (high problem effort, high content effort), Cluster 1 (high problem effort, medium content effort), and Cluster 4 (high problem effort, low content effort). Cluster 0 has the strongest results with 61.08% of students achieving a score of 301-400, and Cluster 1 and 4 having 56% of students in the same score range. Cluster 0 has strong problem and content effort metrics which likely contributed to their higher marks.

The centroids of Cluster 1 and 4 distinguish themselves by their interaction with the slides, with Cluster 1 (high problem effort, medium content effort), showing high levels of slide viewership and low levels of slide completion, whilst Cluster 4 (high problem effort, low content effort) has low metrics in both slide viewership and slide completion. Despite both clusters having 56% of students achieving high marks, Cluster 4 shows 11% of students obtaining a mark of 0-100, while Cluster 1 has 8% of students in this score range. This can indicate that whilst some students in Cluster 4 could have previous knowledge allowing them to score highly without engaging with the course content, this cluster may also include some students who did not engage in the course, and therefore contributed to the lower marks in this cluster.

4.2.4 Comparison of cluster correlation coefficient scores

The correlation coefficient can show the relationship between the cluster features and a student's overall score. The correlation coefficient in Table 4.15 shows how the clusters are correlated with overall score and measure the relationship between the feature variables and score variable. A correlation coefficient of -1 indicates a strong negative relationship, a coefficient of 0 indicates no relationship and a coefficient of 1 indicates a strong positive relationship [33].

TABLE 4.15 CORRELATION COEFFICIENT FOR EACH CLUSTERING FEATURES RELATIONSHIP TO SCORE

Clustering	Correlation Coefficient
FSLSM behaviour	0.393
Procrastination time	0.540
Procrastination effort	0.512

The procrastination time features are most highly correlated with score compared to the other two clusterings. The correlation coefficient for procrastination time features is also similar to the procrastination effort correlation coefficient which may be because of the inclusion of the percentage of completed problems in both groups of features which would indicate a student's overall score, as students with a high submission completion rate would generally also have a high overall score. Therefore, the clusters not only produce meaningful insight into the types of students in the Grok course, but the procrastination features can also assist with predicting a student's score to a medium level of accuracy.

4.2.5 Overall comparison of clusters and scores

Whilst the FSLSM behaviour, procrastination time and procrastination effort clusters show promising ways to measure student learning style, the procrastination clusters provide more correlation and insight between learning style and score. FSLSM behaviour clusters showed a relatively similar distribution of scores between each cluster, and this is supported by literature which states the FSLSM learning styles do not indicate which type will lead to superior performance [8]. Therefore, the procrastination clusters best allow NCSS Grok educators to not only categorise student behaviour but also use these clusters to indicate which students may be struggling with course performance and could benefit from additional teacher supports.

4.3 Objective 3 results: Determine if gender impacts learning style and scores

The gender demographics of each clustering will be investigated to further profile the clusters. The beginners Python NCSS Grok course labelled student gender as male, female or not specified.

The distribution of vectors of each gender for the FSLSM behavioural features is outlined in Table 4.16 and the distribution of vectors for each gender for the procrastination features is in Table 4.17.

TABLE 4.16 DISTRIBUTION OF VECTORS BY GENDER IN FSLSM BEHAVIOURAL DATA

Gender	Total vectors	Percentage of vectors
Female	4982	31.67%
Male	7458	47.40%
None	3293	20.93%
Total	15733	

TABLE 4.17 DISTRIBUTION OF VECTORS BY GENDER IN PROCRASTINATION DATA

Gender	Total vectors	Percentage of vectors
Female	6171	31.29%
Male	9463	47.97%
None	4090	20.74%
Total	19724	

4.3.1 FSLSM behaviour cluster demographics

Overall, the FSLSM behaviour cluster's distribution of scores between genders has a similar spread of low, medium and high scores in Figure 4.4, Figure 4.5 and Figure 4.6, and represented numerically in Table 4.18, 4.19 and 4.20. This indicates that each gender is represented evenly between each score type within each cluster. The one cluster that did have differing results between genders was Cluster 3. Of the males in Cluster 3, 54.70% of them were in the high score range compared to 73.44% of females and 76.86% of non-specified gendered students in the high score range. Students from this cluster had low slide completion and low slide viewership but had a high number of terminal runs and time spent on problems indicating they were engaging in mostly the problem components of the course. Therefore, this implies females and non-specified gendered students with Cluster 3 behaviours were able to achieve more highly than males in this cluster.

In Cluster 1 and 3, males had a higher percentage of upper mid-tier scores in the 201-300 range with around 13-24% of students achieving this score, compared to 8-17% of students who were female or non-disclosed gender in this same score range. Overall, the FSLSM behaviour cluster scores do not show a notable distinction in the scores achieved by different genders in each cluster.

FIGURE 4.4 PERCENTAGE OF FSLSM BEHAVIOURAL CLUSTER MEMBERSHIP FOR FEMALE STUDENTS

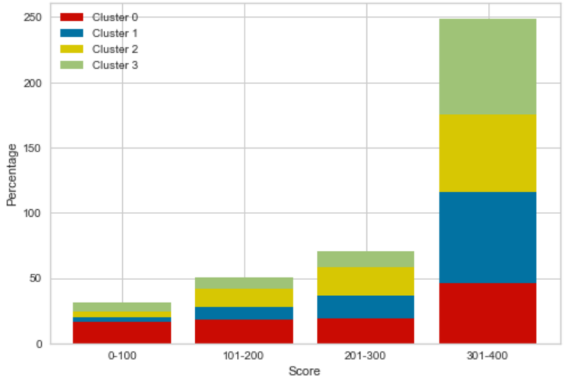


FIGURE 4.5 PERCENTAGE OF FSLSM BEHAVIOURAL CLUSTER MEMBERSHIP FOR MALE STUDENTS

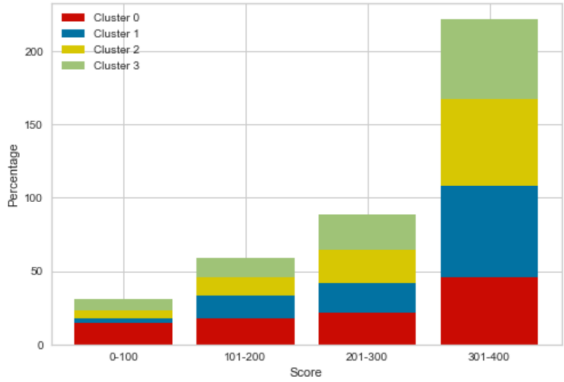


FIGURE 4.6 PERCENTAGE OF FSLSM BEHAVIOUR CLUSTER MEMBERSHIP FOR NON-DISCLOSED GENDER STUDENTS

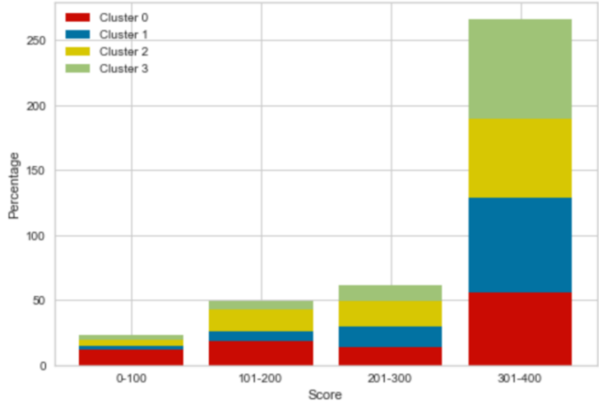


TABLE 4.18 PERCENTAGE OF FSLSM BEHAVIOUR CLUSTER VECTORS FOR FEMALES IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	901	16.65%	17.87%	19.09%	46.39%
Cluster 1	982	3.36%	9.47%	17.62%	69.55%
Cluster 2	2907	4.64%	14.59%	21.60%	59.17%
Cluster 3	192	6.25%	8.33%	11.98%	73.44%

TABLE 4.19 PERCENTAGE OF FSLSM BEHAVIOUR CLUSTER VECTORS FOR MALES IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	1986	14.25%	17.98%	21.70%	46.07%
Cluster 1	1114	3.14%	15.35%	20.02%	61.49%
Cluster 2	3901	5.56%	12.43%	22.82%	59.19%
Cluster 3	457	7.88%	13.13%	24.29%	54.70%

TABLE 4.20 PERCENTAGE OF FSLSM BEHAVIOUR CLUSTER VECTORS FOR NON-DISCLOSED GENDER IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	595	11.93%	18.49%	13.95%	55.63%
Cluster 1	663	2.87%	7.84%	15.69%	73.60%
Cluster 2	1793	4.24%	16.40%	19.35%	60.01%
Cluster 3	242	4.54%	6.20%	12.40%	76.86%

The distribution of each gender in each behavioural cluster in Table 4.21 shows there is a comparatively high number of males in Cluster 0 compared to females and non-disclosed gender, but a low percentage of males in Cluster 1. This distribution indicates that males have lower slide completion rates. This is because there is a high percentage of males in Cluster 0 which on average has a slide completion rate of 24.6%, whilst there are a low number of males in Cluster 1 which has high slide completion rates of 84.9%. as low slide completion rates.

TABLE 4.21 PERCENTAGE DISTRIBUTION OF FEMALE, MALE AND NON-DISCLOSED GENDERED STUDENTS IN EACH FSLSM BEHAVIOURAL CLUSTER

	Female		**Male**		**Non-disclosed**	
	Number	*Percentage*	*Number*	*Percentage*	*Number*	*Percentage*
Cluster 0	901	18.10%	1986	26.65%	595	18.10%
Cluster 1	982	19.70%	1114	14.93%	663	20.13%
Cluster 2	2907	58.35%	3901	52.29%	1793	54.42%
Cluster 3	192	3.85%	457	6.13%	242	7.35%
Total	4982		7458		3293	

Overall, there is a consistent distribution of genders between each FSLSM behavioural cluster and all genders were represented within each score range in these clusters.

4.3.2 Procrastination time cluster demographics

Similarly, to the FSLSM clusters, the gender distribution is consistent between all the procrastination time clusters as evident in Figure 4.7, 4.8 and 4.9, and numerically represented in Tables 4.22, 4.23, 4.24.

FIGURE 4.7 PERCENTAGE OF PROCRASTINATION TIME CLUSTER MEMBERSHIP FOR FEMALE STUDENTS

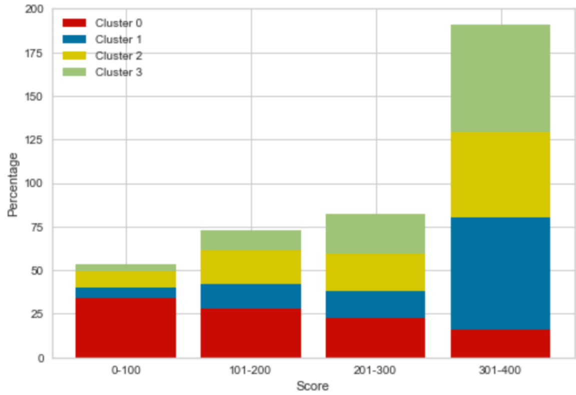


FIGURE 4.8 PERCENTAGE OF PROCRASTINATION TIME CLUSTER MEMBERSHIP FOR MALE STUDENTS

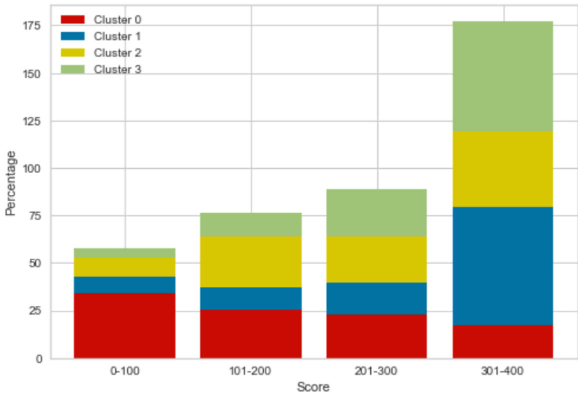


FIGURE 4.9 PERCENTAGE OF PROCRASTINATION TIME CLUSTER MEMBERSHIP FOR NON-DISCLOSED GENDERED STUDENTS

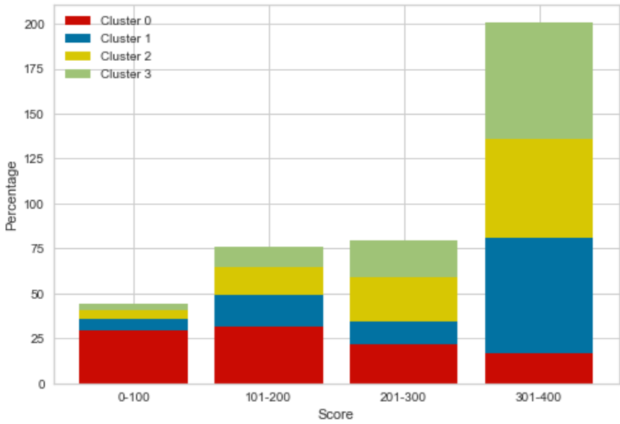


TABLE 4.22 PERCENTAGE OF PROCRASTINATION TIME CLUSTER VECTORS FOR FEMALES IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	965	33.99%	27.77%	22.59%	15.65%
Cluster 1	2626	5.75%	14.05%	15.39%	64.81%
Cluster 2	927	9.71%	19.96%	21.36%	48.97%
Cluster 3	1653	4.24%	11.31%	23.17%	61.28%

TABLE 4.23 PERCENTAGE OF PROCRASTINATION TIME CLUSTER VECTORS FOR MALES IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	1637	34.39%	25.54%	23.03%	17.04%
Cluster 1	4047	8.58%	11.81%	16.90%	62.71%
Cluster 2	1310	9.62%	26.64%	23.89%	39.85%
Cluster 3	2469	5.10%	12.15%	25.24%	57.51%

TABLE 4.24 PERCENTAGE OF PROCRASTINATION TIME CLUSTER VECTORS FOR NON-DISCLOSED GENDER IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	568	29.23%	31.87%	21.83%	17.07%
Cluster 1	2033	6.39%	17.46%	12.64%	63.51%
Cluster 2	480	5.00%	15.21%	24.37%	55.42%
Cluster 3	1009	3.57%	11.10%	20.61%	64.72%

Males have the lowest percentage of scores in the high range compared to females and non-disclosed in Cluster 2 (low procrastination). Of the males in Cluster 2, 39.85% achieved high scores, compared to 49.97% of female and 55.42% of non-disclosed gender students. Cluster 2 is characterised by spending longer on problems and this score distribution can indicate that male students may procrastinate or be disengaged with the problems while working on them compared to other students, resulting in a lower percentage of students with a high score within the low procrastination cluster.

TABLE 4.25 PERCENTAGE DISTRIBUTION OF FEMALE, MALE AND NON-DISCLOSED GENDERED STUDENTS IN EACH PROCRASTINATION TIME CLUSTER

	Female		**Male**		**None**	
	Number	*Percentage*	*Number*	*Percentage*	*Number*	*Percentage*
Cluster 0	965	15.64%	1637	17.30%	568	13.88%
Cluster 1	2626	42.55%	4047	42.76%	2033	49.71%
Cluster 2	927	15.02%	1310	13.85%	480	11.74%
Cluster 3	1653	26.79%	2469	26.09%	1009	24.67%
Total	6171		9463		4090	

Additionally, males were slightly more represented in Cluster 0 (high procrastination) compared to their peers as seen in Table 4.25, with 17.30% of males in this cluster compared to 15.64% of females. Overall, the distribution of students within each cluster is similar based on gender and highlights students of all genders can have both low, medium and high procrastination rates throughout the course.

4.3.3 Procrastination effort cluster demographics

Procrastination effort clusters have a similar spread of genders in each score range within each cluster. Results for procrastination effort clusters are in Figure 4.10, 4.11, 4.12, and Table 4.26, 4.27, 4.28. It is notable that in Cluster 4 (high problem effort, low content effort) there is low male representation in the high score range, with only 50.65% of males obtaining high scores compared to 60.37% of females and 68.36% of non-disclosed students.

FIGURE 4.10 PERCENTAGE OF PROCRASTINATION EFFORT CLUSTER MEMBERSHIP FOR FEMALE STUDENTS

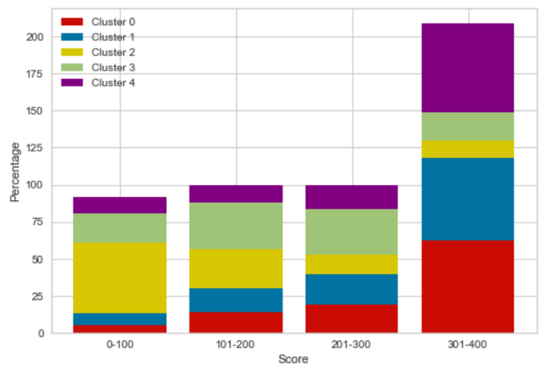


FIGURE 4.11 PERCENTAGE OF PROCRASTINATION EFFORT CLUSTER MEMBERSHIP FOR MALE STUDENTS

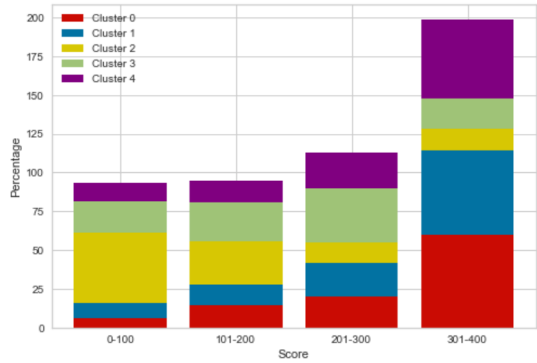


FIGURE 4.12 PERCENTAGE OF PROCRASTINATION EFFORT CLUSTER MEMBERSHIP FOR NON-DISCLOSED GENDERED STUDENTS

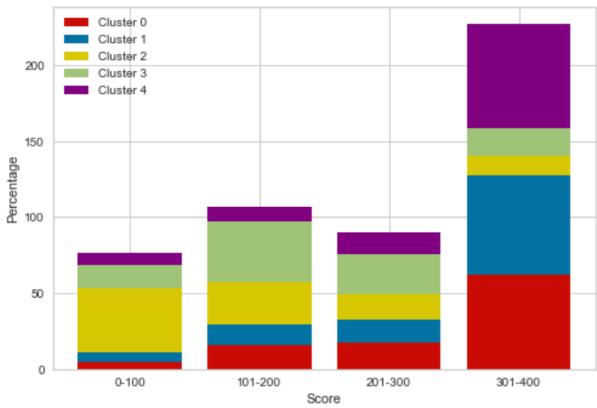


TABLE 4.26 PERCENTAGE OF PROCRASTINATION EFFORT CLUSTER VECTORS FOR FEMALES IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	4105	4.75%	13.76%	19.15%	62.34%
Cluster 1	682	8.07%	16.28%	20.23%	55.42%
Cluster 2	528	48.29%	26.33%	13.07%	12.31%
Cluster 3	475	19.37%	31.37%	30.73%	18.53%
Cluster 4	381	11.02%	11.81%	16.80%	60.37%

TABLE 4.27 PERCENTAGE OF PROCRASTINATION EFFORT CLUSTER VECTORS FOR MALES IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	5374	5.82%	14.26%	20.30%	59.62%
Cluster 1	1415	10.32%	13.43%	21.70%	54.55%
Cluster 2	979	44.94%	27.99%	13.18%	13.89%
Cluster 3	688	20.35%	25.00%	34.74%	19.91%
Cluster 4	1007	12.21%	14.20%	22.94%	50.65%

TABLE 4.28 PERCENTAGE OF PROCRASTINATION EFFORT CLUSTER VECTORS FOR NON-DISCLOSED GENDER IN EACH SCORE RANGE

		Percentage of vectors within each score range			
	Total vectors	Score 0-100	Score 101-200	Score 201-300	Score 301-400
Cluster 0	2656	4.78%	15.96%	17.17%	62.09%
Cluster 1	499	6.01%	13.63%	15.23%	65.13%
Cluster 2	301	42.19%	27.58%	16.94%	13.29%
Cluster 3	280	15.72%	39.64%	26.43%	18.21%
Cluster 4	354	7.91%	9.89%	13.84%	68.36%

Whilst the distribution of clusters in each gender is similar in Table 4.29, there is some overrepresentation of males in Cluster 1 (high problem effort, medium content effort), Cluster 2 (low problem effort, low content effort) and Cluster 4 (high problem effort, low content effort) compared to the females and non-disclosed gender. Furthermore, males are underrepresented in Cluster 0 (high problem effort, high content effort) where 56.79% of males are in this cluster compared to 66.52% of females and 64.94% of non-disclosed. This reinforces the finding that male students are less engaged with the course content compared to other genders.

TABLE 4.29 PERCENTAGE DISTRIBUTION OF FEMALE, MALE AND NON-DISCLOSED GENDERED STUDENTS IN EACH PROCRASTINATION EFFORT CLUSTER

	Female		**Male**		**None**	
	Number	*Percentage*	*Number*	*Percentage*	*Number*	*Percentage*
Cluster 0	4105	66.52%	5374	56.79%	2656	64.94%
Cluster 1	682	11.06%	1415	14.95%	499	12.20%
Cluster 2	528	8.55%	979	10.35%	301	7.35%
Cluster 3	475	7.70%	688	7.27%	280	6.85%
Cluster 4	381	6.17%	1007	10.64%	354	8.66%
Total	6171		9463		4090	

4.4 Objective 4 results: Investigate how student learning style changes over time and impacts drop off rates

Student learning style is dynamic and can change over time throughout the course duration. Markov chains were used to analyse how student learning style changed by measuring the percentage of transitions between clusters each week. The aggregated versions of the Markov chains are in Figure 4.13, 4.14 and 4.15, which shows the students from Week 1 and the percentage transitions of these students to clusters in the proceeding weeks. In addition to the clusters generated in Section 4, two other clusters were added to the Markov chain calculations. One cluster is No Attempt which means a student did not have any course activity for the week. In other words, the vector for that week would be all NaN. The second cluster is Partial Attempt which means students were engaged in some, but not all features measured. The Markov chain analysis aims to help indicate to teachers when a student is at a high risk of dropping out of the course and when to best engage students to prevent drop-out rates.

4.4.1 FSLSM behaviour Markov Chain

FSLSM behaviour Markov chains are visualised as a directed graph in Figure 4.13 and as a series of pie charts in Figure 4.14.

FIGURE 4.13 FSLSM BEHAVIOURAL MARKOV CHAIN. PERCENTAGE OF TRANSITIONS FROM WEEK 1 CLUSTERS TO CLUSTERS IN PROCEEDING WEEKS

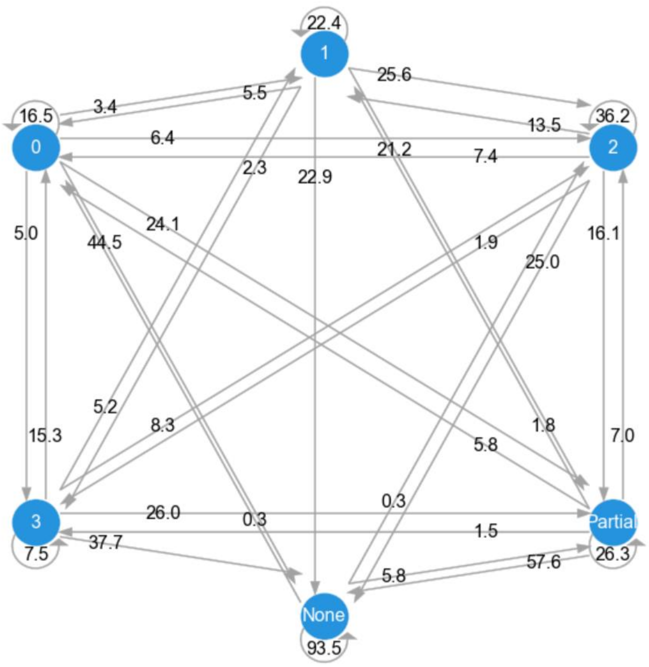
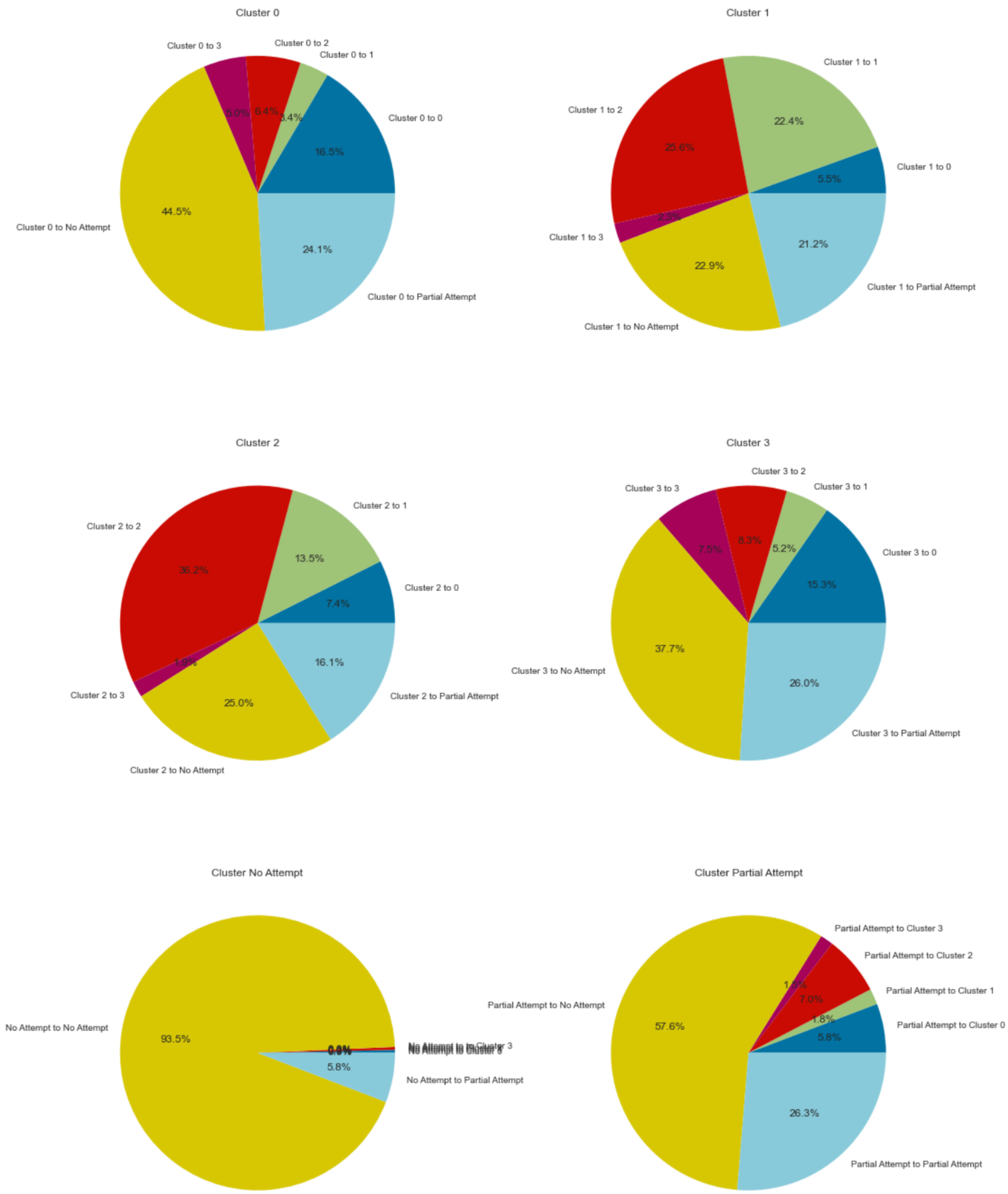


FIGURE 4.14 FSLSM BEHAVIOURAL MARKOV CHAIN REPRESENTED AS PIE CHARTS. PERCENTAGE OF TRANSITIONS FROM WEEK 1 CLUSTERS TO CLUSTERS IN PROCEEDING WEEKS



4.4.1.1 Transitions from procrastination time clusters to No Attempt

Figure 4.13 and 4.14 shows the Markov chain for the FSLSM behaviour clusters. The Cluster 0 nodes show that 44.5% of vectors in Cluster 0 in Week 1 transitioned to No Attempt in any of the proceeding weeks. This is the highest percentage transition to No Attempt of any of the FSLSM clusters. This indicates that Cluster 0 students are of high risk of dropout in the course.

Cluster 3 also had a high number of transitions to No Attempt, with 37.7% of students in Cluster 1 in Week 1, transitioning to No Attempt in any of the proceeding weeks. Both Cluster 0 and Cluster 3 show low levels of slide completion rates of 24% slides completed on average. Therefore, low slide completion rates can indicate low levels of student course engagement and can be an indicator of future disengagement with the course.

4.4.1.2 Transitions from a FSLSM cluster to the same cluster

Cluster 2 had the highest rates of cluster stagnation of 36.2%. Each cluster except Cluster 2 show similar rates of movement to other FSLSM clusters (not inclusive of No Attempt or Partial Attempt), compared to remaining in the existing cluster. These results are surprising and differ from previous work completed by McBroom et al. [12] which indicate learning styles of students largely remain the same. The results of this thesis instead show that majority of behavioural groupings of students transition to another cluster at some point in the course.

4.4.1.3 Transitions between FSLSM clusters

Cluster 1 and 2 shows high movement between Cluster 1 and 2. These two clusters are similar, and both have high slide completion rates of 83% on average, except Cluster 1 spends more time on problems than Cluster 2. The relatively high number of transitions between Cluster 1 and 2 suggests students can switch between behaviour types each week, perhaps as question difficulty increases.

Conversely, there is only moderate movement between Cluster 0 and Cluster 3, despite those clusters sharing similar centroid with low slide completion rates. Cluster 3 does show a 15.3% transition rate to Cluster 0 but this is not reciprocated from Cluster 0 to Cluster 3 with this rate being only 5.0%. The transitions observed from Cluster 3 to Cluster 0 are supported by the previous findings in this thesis that Cluster 3 is one of the highest performing clusters with 64.76% of students obtaining a score within the range 301-400, while only 47.79% of students from Cluster 0 are able to achieve the same mark. The movement from Cluster 3 to Cluster 0 indicates that students who are initially performing highly with low slide completion rates, then transition to Cluster 0 and perform less highly in proceeding weeks. The cause of this lower performance may be due to students creating a habit of slide non-completion which does not sustain them during the more difficult weeks of the course or there may be other external factors impacting student performance in this cluster.

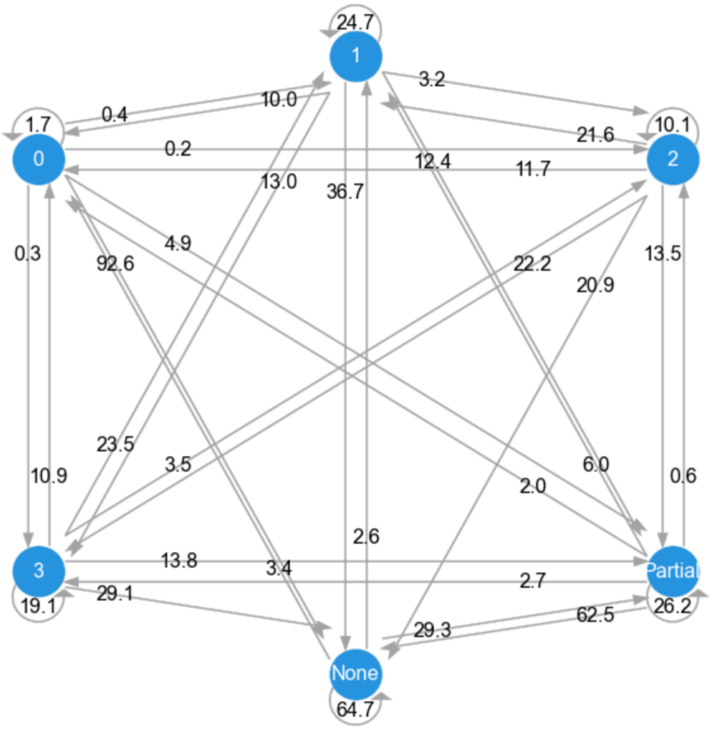
4.4.1.4 Transitions from No Attempt and Partial Attempt to FSLSM clusters

Overall, once a student is in the No Attempt cluster they do not re-engage with the course in following weeks as 93.5% of students remain in this cluster. This is similar for Partial Attempt where 26.3% students will either stay as partially attempting the course and 57.6% of students transition to No Attempt. From the Partial Attempt cluster 7.0% of students transition to Cluster 2 and 5.8% transition to Cluster 0. These are the two lowest performing clusters with only 47.79% of students in Cluster 0 and 59.35% of students in Cluster 2 performing in the highest score range. Therefore, whilst some students are observed transitioning from the Partial Attempt cluster to the FSLSM clusters, these FSLSM clusters have a significant lower performance. This can suggest that consistent engagement with the course is a factor that can impact a student's overall course performance.

4.4.2 Procrastination time Markov Chain

Results for procrastination time Markov chains are displayed in Figure 4.15 as a directed graph and in Figure 4.16 as pie charts.

FIGURE 4.15 PROCRASTINATION TIME MARKOV CHAIN. PERCENTAGE OF TRANSITIONS FROM WEEK 1 CLUSTERS TO CLUSTERS IN PROCEEDING WEEKS



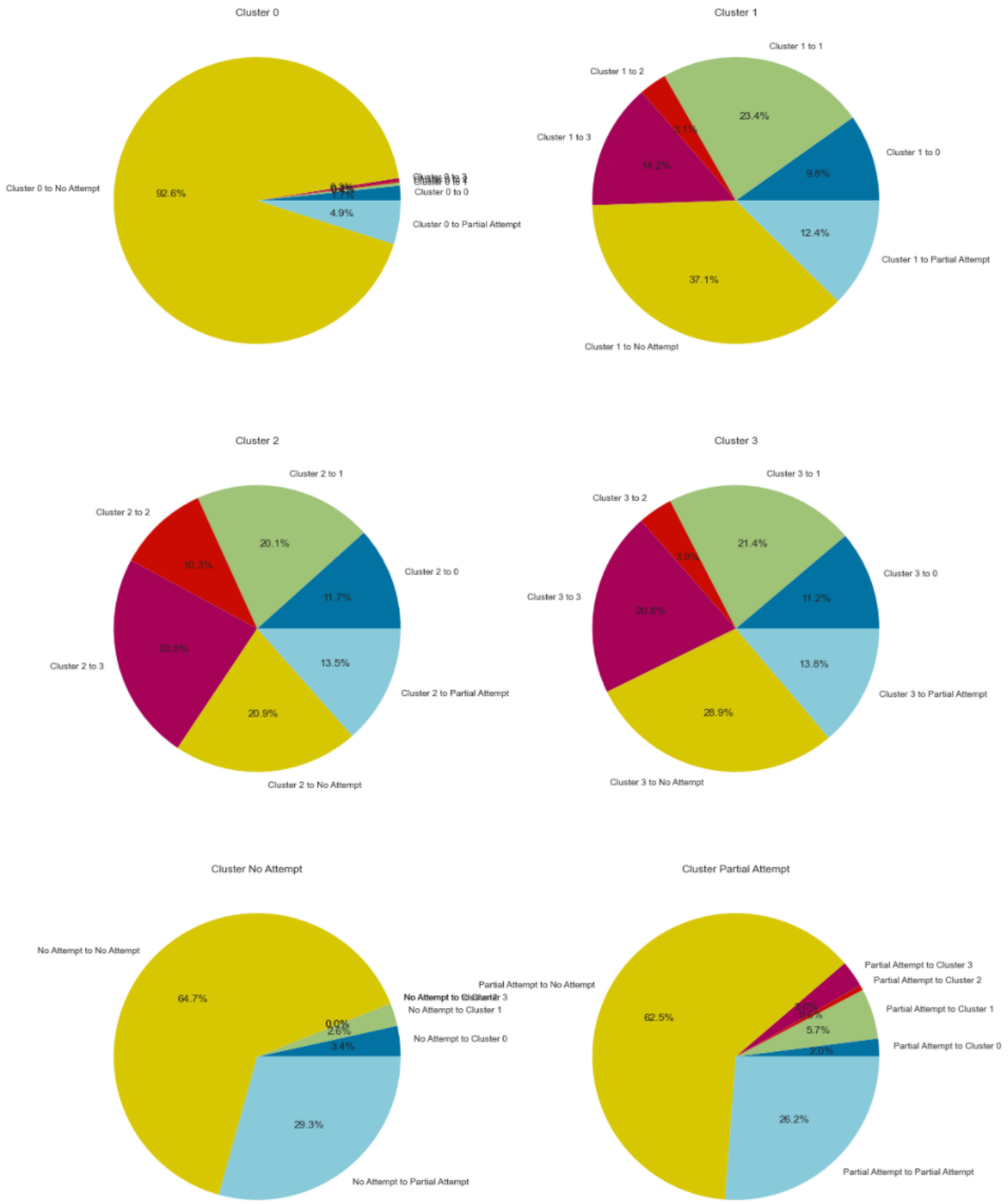
4.4.2.1 Transitions from procrastination time clusters to No Attempt

The procrastination time Markov chains in Figure 4.15 and 4.16, show a particularly high transition to No Attempt for Cluster 0 (high procrastination) with 92.6% of students who were in Cluster 0 in Week 1 moving to the No Attempt cluster in the proceeding weeks. This is higher than the percentage of students who began in the No Attempt cluster, of which 64.7% of those students remained in the No Attempt clusters. The high number of Cluster 0 to No Attempt transitions highlights that this cluster is of very high risk of disengaging with the course.

4.4.2.2 Transitions from procrastination time clusters to the same cluster

Similarly, to the FSLSM behaviour clusters, the procrastination time Markov chains indicate there are a higher percentage of students who transition to another cluster than those students who remain in the same cluster. Cluster 1 is the procrastination time cluster with the highest percentage of students remaining in the cluster with a 23.4% retention rate.

FIGURE 4.16 PROCRASTINATION TIME MARKOV CHAIN REPRESENTED AS PIE CHARTS. PERCENTAGE OF TRANSITIONS FROM WEEK 1 CLUSTERS TO CLUSTERS IN PROCEEDING WEEKS



4.4.2.3 Transitions between procrastination time clusters

Cluster 2 (low procrastination) and Cluster 3 (moderate procrastination) show a relatively smaller number of drop off rates but a high number of transitions to other clusters. Both Cluster 2 and 3 are similar, where students start questions early shortly after they are released, but Cluster 2 spends almost double the average time on content than Cluster 3. The interchangeableness of these clusters means students may be spending more or less time on content depending on their skill level and the difficulty of the content for the week.

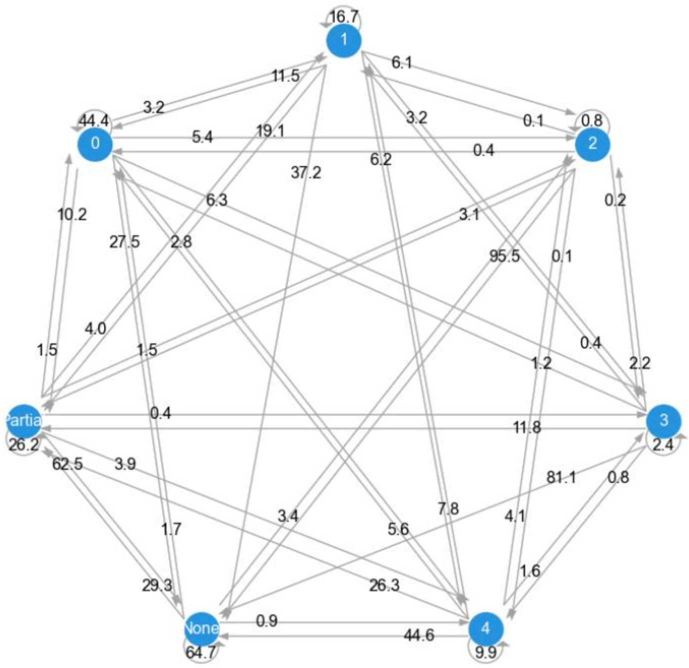
4.4.2.4 Transitions from No Attempt and Partial Attempt to procrastination time clusters

Similarly, to the FSLSM behaviour cluster transitions, once a student is in the No Attempt and Partial Attempt clusters there is only a small trend of moving to any of the procrastination time clusters. However, Cluster 1 showed the highest percentage of transitions from Partial Attempt, with a 5.7% transition rate, where Cluster 1 was the low procrastination group with the highest percentage of scores in the high score range. The notable number of transitions to Cluster 1 may be because this cluster comprised of 41.48% of all vectors in this clustering and therefore also represented a larger percentage of vectors in this transition.

4.4.3 Procrastination effort Markov Chain

Procrastination effort Markov chain results are shown in Figure 4.17 and Figure 4.18.

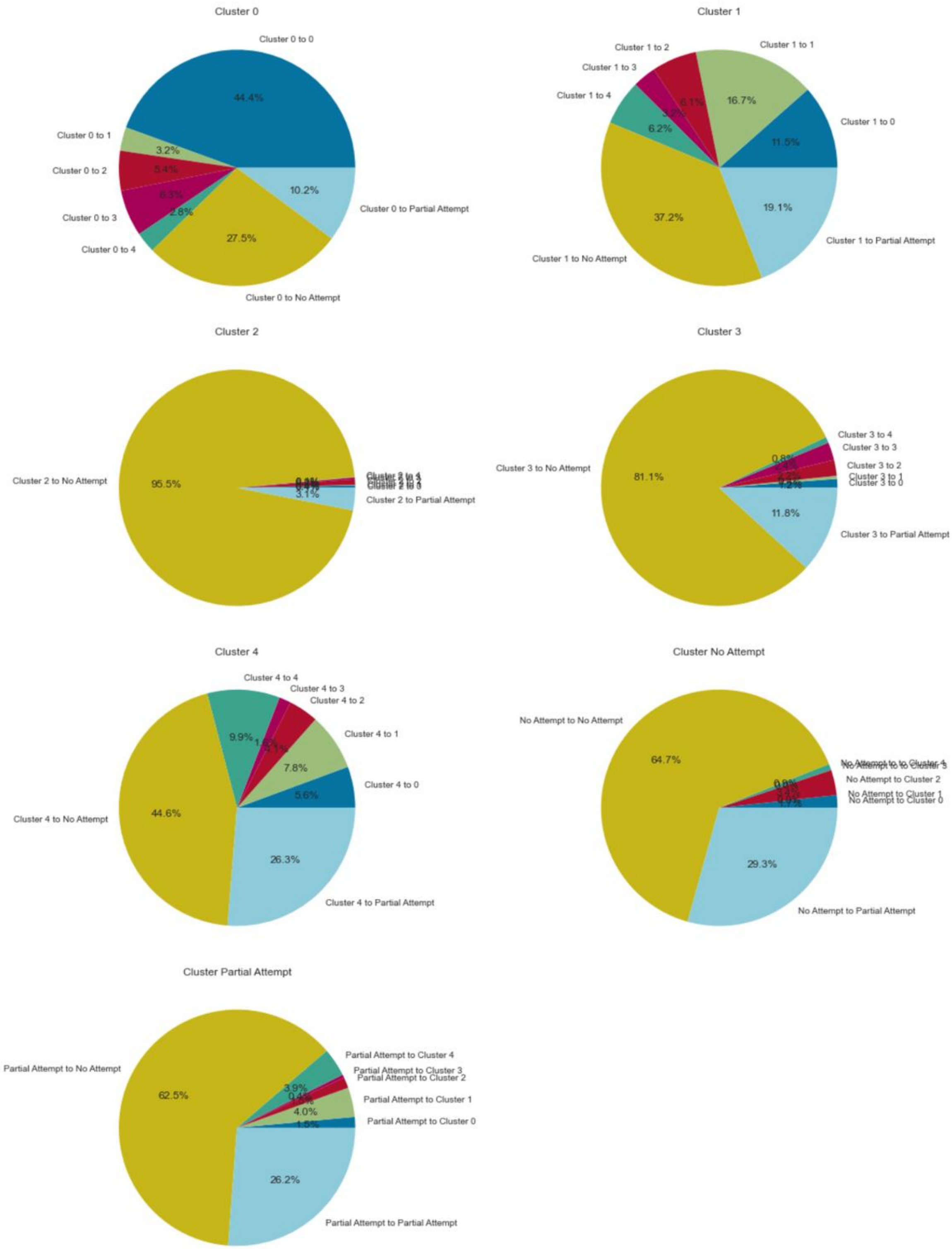
FIGURE 4.17 PROCRASTINATION EFFORT MARKOV CHAIN. PERCENTAGE TRANSITION OF CLUSTERS FROM WEEK 1 TO CLUSTERS IN PROCEEDING WEEKS



4.4.3.1 Transitions from procrastination effort clusters to No Attempt

Cluster 2 (low problem effort, low content effort) and Cluster 3 (low problem effort, high content effort) consists of students with the highest risk of dropout with 95.5% and 81.1% of students who began in this cluster moving to No Attempt in previous weeks respectively. Alarmingly, this percentage is higher than the students who began in the No Attempt group in Week 1. Whilst this could be expected for Cluster 2, which had low activity in all effort features measured, Cluster 3 had 72% of slides viewed on average, which would normally indicate high engagement with the content. Therefore, for the procrastination effort cluster transitions, the Markov chains highlighted at-risk students who may have gone unnoticed due to high measures on some effort features.

FIGURE 4.18 PROCRASTINATION EFFORT MARKOV CHAIN REPRESENTED AS PIE CHARTS. PERCENTAGE TRANSITIONS OF CLUSTERS FROM WEEK 1 TO CLUSTERS IN PROCEEDING WEEKS



4.4.3.2 Transitions from procrastination effort clusters to the same cluster

The highest number of same cluster transitions is in Cluster 0 where 44.4% of students in Cluster 0 in Week 1 remained in this cluster. Cluster 0 represented the high problem effort, high content effort students who were also highly represented in the high score ranges. This shows Cluster 0 students not only display commitment to engaging with problems and slide content in the course, they also maintain their study of both problem and slide course components beyond Week 1 of the course.

4.4.3.3 Transitions between procrastination effort clusters

There are transitions between Cluster 1 to Cluster 0 where 11.5% of students in Cluster 1 move to Cluster 0 at some point in the course after Week 1. The two clusters have similar characteristics of Cluster 0 being high problem effort, high content effort, and Cluster 1 being high problem effort, medium content effort clusters. Therefore, students may vary their level of content engagement in the course throughout the course duration.

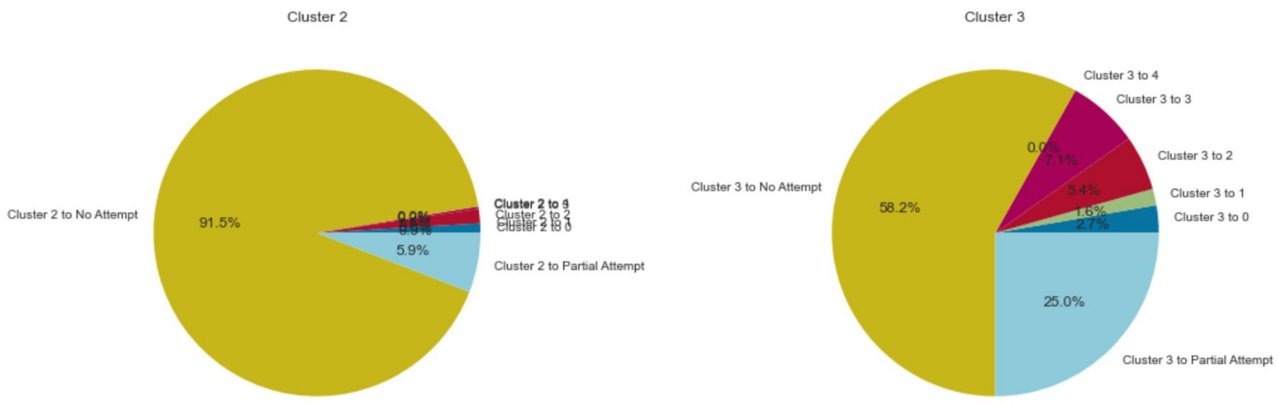
4.4.3.4 Transitions from No Attempt and Partial Attempt to procrastination effort clusters

Overall, each Markov chain shows a pattern that once a student has made no attempt for the week, they are unlikely to re-engage with the course. Similarly, to the procrastination time clusters the procrastination effort Markov chain shows there is substantial movement between No Attempt and Partial Attempt groupings.

4.4.4 Changes in student learning between consecutive weeks

The Markov chains can also provide more detailed insights when calculated between each consecutive week and a full visualisation of all consecutive Markov chains are in Appendix B, C and D. An example of a consecutive week Markov chain for the procrastination effort features is shown in Figure 4.19. Previously in Figure 4.17 and Figure 4.18, it was identified that Cluster 2 (low problem effort, low content effort) and Cluster 3 (low problem effort, high content effort) are the highest risk students of disengaging with the course. By using a consecutive week Markov chain we can determine not only which students are most at-risk, but when they are most at-risk of disengaging with the course. For example, when looking at procrastination effort transitions between Week 1 and Week 2 in Figure 4.19, we can see that Cluster 2 had a 91.5% drop off rate after the first week. Therefore, Cluster 2 students would require intervention as early as during the first week of the course to prevent their course engagement levels from decreasing.

FIGURE 4.19 PROCRASTINATION EFFORT MARKOV CHAIN FOR CLUSTER 2 AND 3 FROM WEEK 1 TO WEEK 2



Conversely Cluster 3 has a lower percentage of transitions to No Attempt from Week 1 to Week 2. This implies Cluster 3's transitions to No Attempt were gradual and spread across multiple weeks whilst Cluster 2's drop-off rates are more sudden, occurring as early as in Week 1. Therefore, the Markov chains can inform educators on which students are of highest risk of disengaging from the course as well as highlighting when these students are most risk of course disengagement.

The FSLSM behaviour, procrastination time and procrastination effort cluster movement to Partial Attempt in Figure 4.20, 4.21 and 4.22, all show a drop in Partial Attempt transitions in Week 4 followed by a rise in Week 5. The drop in Partial Attempts in Week 4 may reflect student's ability to learn from the course content and improve in previous weeks. The drop in Partial Attempt may also indicate that students found Week 4 easier and were able to complete more of the content and questions. It should be noted that the students who saw an increase in Partial Attempts transitions in Week 4 also had a drop in No Attempt in the same week. Therefore, we can assume these students who were previously not attempting questions from these clusters were able to attempt questions in Week 4 due to the easier content. However, the rise in Week 5 could indicate that Week 5 tasks were particularly challenging, or students were disengaging from the course towards the final week.

FIGURE 4.200 BEHAVIOUR CLUSTER PERCENTAGE TRANSITIONS BETWEEN CONSECUTIVE WEEKS FROM CLUSTERS TO PARTIAL ATTEMPT

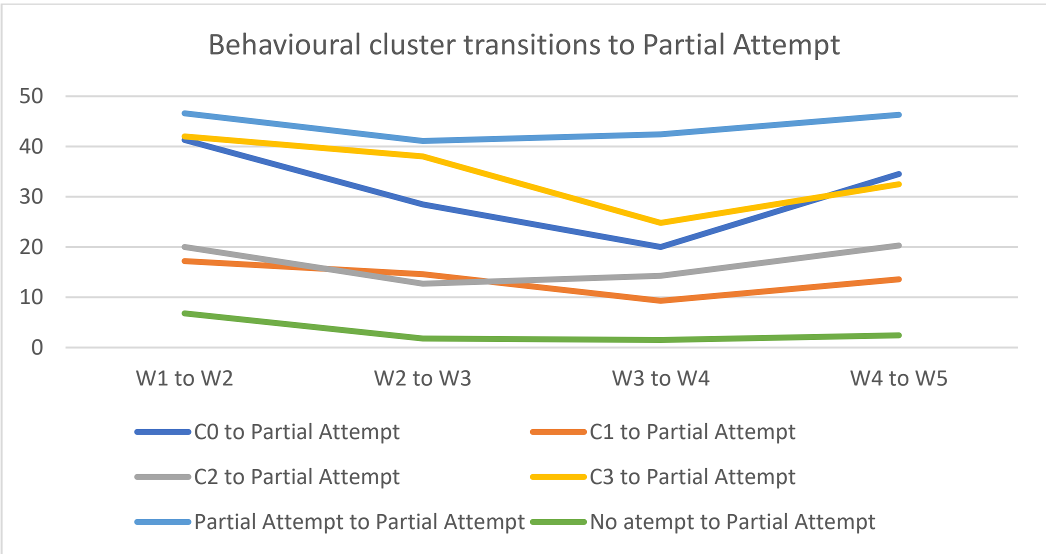


FIGURE 4.211 PROCRASTINATION TIME CLUSTER PERCENTAGE TRANSITIONS BETWEEN CONSECUTIVE WEEKS FROM CLUSTERS TO PARTIAL ATTEMPT

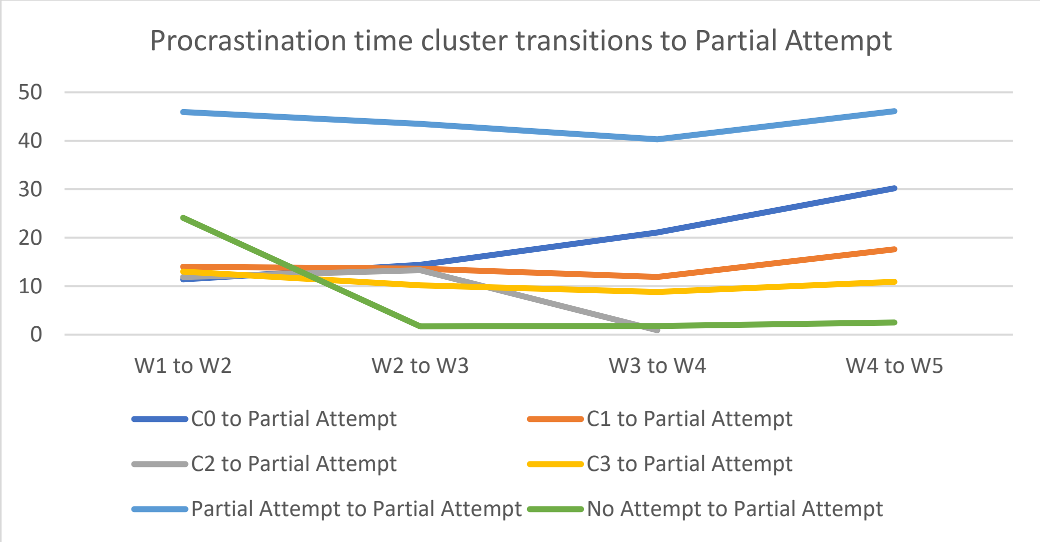
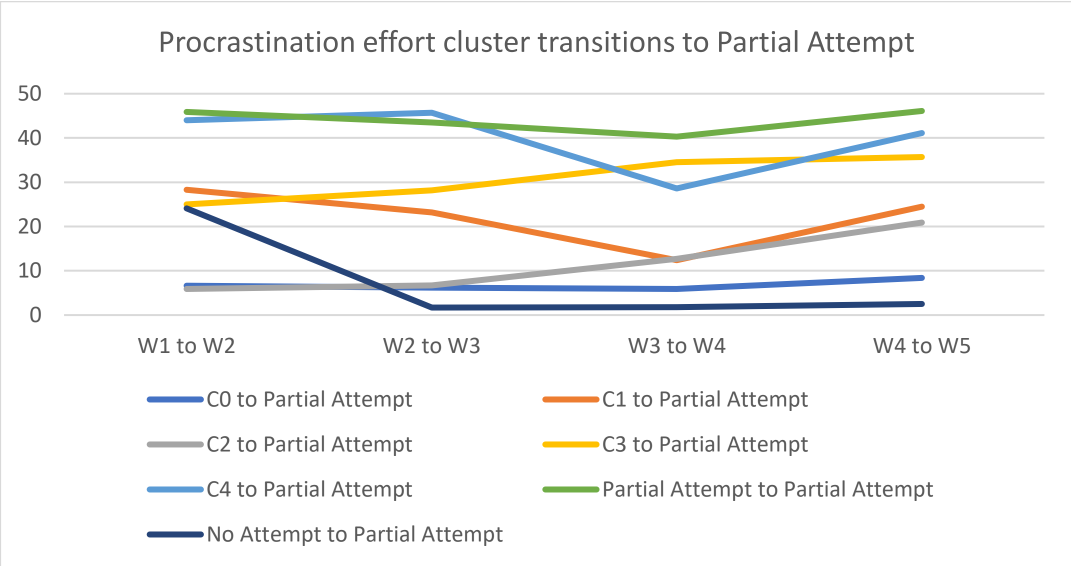


FIGURE 4.222 PROCRASTINATION EFFORT CLUSTER PERCENTAGE TRANSITIONS BETWEEN CONSECUTIVE WEEKS FROM CLUSTERS TO PARTIAL ATTEMPT



Conversely, the rates of No Attempt are relatively consistent within each cluster, as shown in Figure 4.23, 4.24 and 4.25. Each clustering shows that after Week 2, there are almost 100% retention rates in the No Attempt cluster. Additionally, the clusters that had a reduced percentage of No Attempt clusters in Week 5, also had an increased rate of Partial Attempt representation for that week. As question difficulty increases some clusters such as procrastination time Cluster 0, may be more prone to partial course completion rather than completely disengaging from the course.

Therefore, the line graphs representing Markov chain transitions to Partial Attempt and No Attempt highlight the overall trends in all clusters and show how increases in course disengagement tendencies can inform educators on question difficulty and student confidence throughout the course.

FIGURE 4.233 BEHAVIOUR CLUSTER PERCENTAGE TRANSITIONS BETWEEN CONSECUTIVE WEEKS FROM CLUSTERS TO NO ATTEMPT

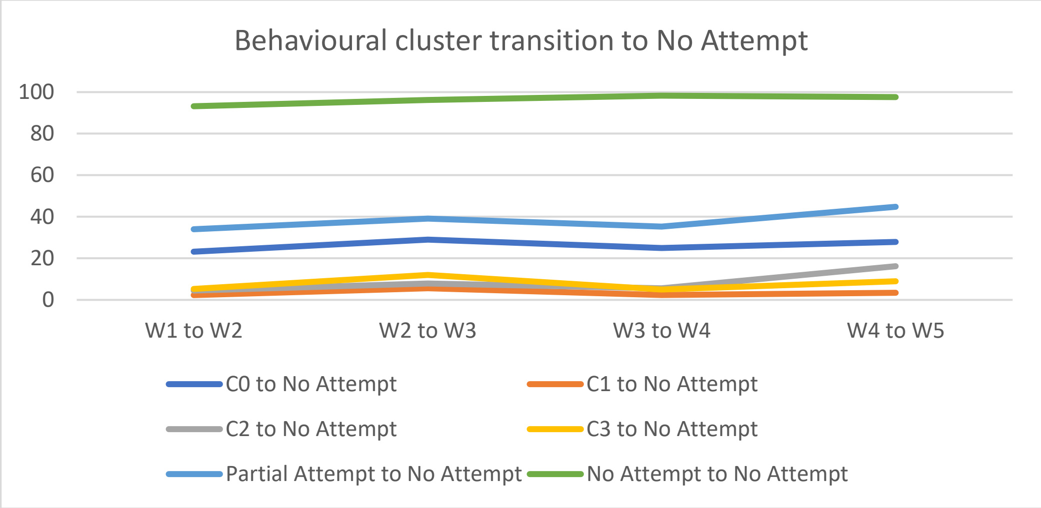


FIGURE 4.24 PROCRASTINATION TIME CLUSTER PERCENTAGE TRANSITIONS BETWEEN CONSECUTIVE WEEKS FROM CLUSTERS TO NO ATTEMPT

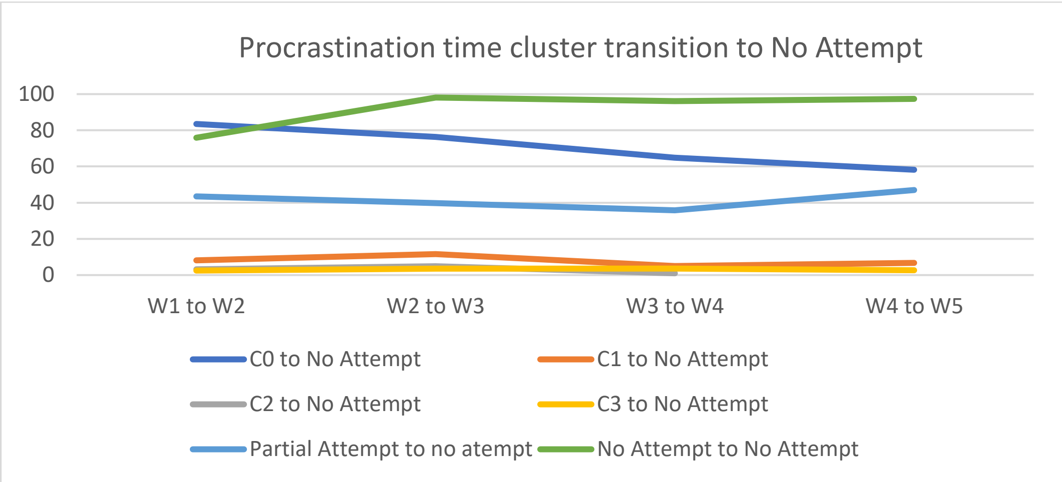
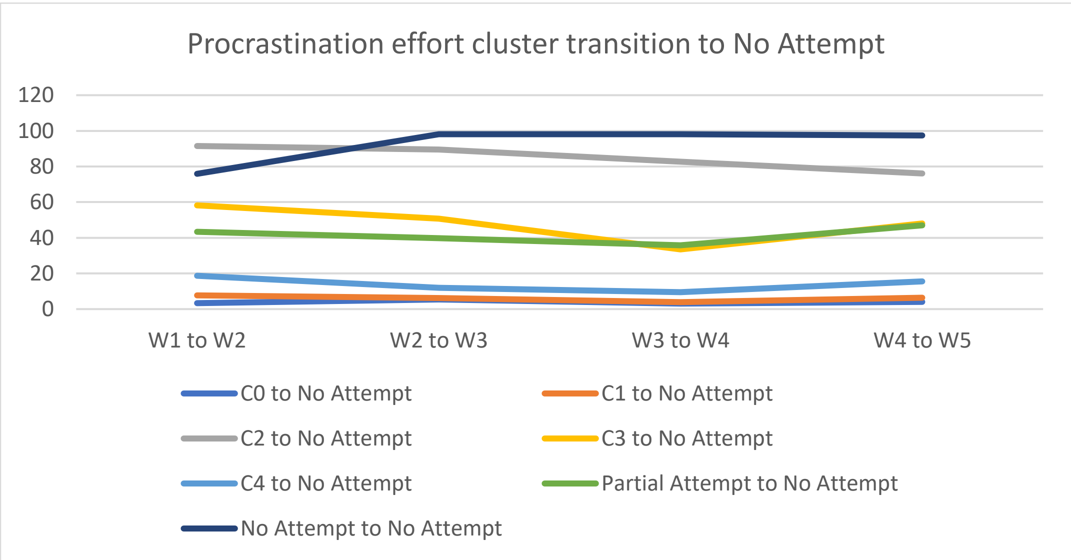


FIGURE 4.255 PROCRASTINATION EFFORT CLUSTER PERCENTAGE TRANSITIONS BETWEEN CONSECUTIVE WEEKS FROM CLUSTERS TO NO ATTEMPT



Chapter 5 Conclusion

The thesis addressed the key objectives and presented a number of contributions in the educational data mining field.

The first objective concerned with investigating the similarities in student learning style based on behaviour and procrastination tendencies, resulted in the contribution of defining behavioural and procrastination features based on Felder-Silverman's theory [9] and Cerezo et al.'s [11] research.

Objective 2 regarding student performance in each cluster, obtained the following results. For the FSLSM features, four clusters were generated which reflected students of different learning styles where all clusters had a similar distribution of low, medium and high scores. This highlights that students of different behavioural learning styles were able to achieve high scores, and performance was not correlated to FSLSM learning behaviours. Conversely, the procrastination features were divided into two groups for clustering, the first group relating to time features and the second group relating to effort features. The time clusters identified one high, one moderate and two low levels of procrastination groups. Clustering of procrastination effort features formulated five different clusters relating to student's problem and slide content completion rates. The effort procrastination clusters had a more distinct distribution of scores between each cluster with the 'low procrastination' and the 'high problem effort, high content effort' clusters representing the students with the highest scores in the 301-400 range. Therefore, this thesis contributed to the research on how to identify struggling students and showed that whilst the FSLSM behaviour clusters can identify different learning behaviours, the procrastination time and effort clusters provide greater insight into the impact of learning style on course performance and score.

A demographic profiling of each cluster answered objective 3, and it was discovered that male students typically were represented by clusters with less slide engagement in both the FSLSM behavioural clusters, procrastination time clusters and procrastination effort clusters. Therefore, the thesis contributed to insight into gendered performance and showed that the female students performed as highly and for some clusters, more highly than males in the course.

The Markov chain results allowed for insight into objective 4, where the student drop-out rates in each set of clusters can identify at-risk students throughout the course duration. Our results showed that students with low procrastination effort features were at high risk of having no attempt in the course after Week 1. Using Markov chains to compare the student's cluster transition between consecutive weeks can highlight when a group of students is most at-risk of dropping out of the course. It was shown that procrastination clusters with low levels of content and slide effort had student's participation rates drop off as early as after Week 1. Therefore, the Markov chain contribution from this research, allows for better information to teachers on when to provide assistance to at-risk students to prevent drop off rates as it was noted that in all clusters, once a student disengages in the course for a week it is unlikely they will reengage in any proceeding weeks.

Overall, our method allows teachers to identify groups of poorly performing students based on their learning styles and can highlight when these students are most at-risk of becoming disengaged from the course. These insights can assist teachers in targeting students who require remedial action and improve student's learning outcomes in the course.

5.1 Future work

Areas of future work to be explored are as follows:

- Additional clustering algorithms can be explored in future research. K-means was the selected clustering algorithm used on the dataset and other clustering techniques such as density-based clustering would allow for a comparison between the spherical clusterings formed by K-means to the arbitrary shaped clusters in density-based clustering. Experimenting with other clustering techniques can allow for a comparison of the Silhouette score and Davies Bouldin score to determine which algorithm creates clusters with strong inter and intra cluster similarity on the dataset.
- Whilst the thesis investigated the demographic profiling of the clusters based on gender, future research can extend the demographic profiling of the clusters to include age or school district. The existing dataset included a range of student year groups from Year 2 to Year 12 and would provide an interesting comparison on how students from different year groups performed in the beginners NCSS Python course.
- Future research can use the features and clusters identified in this thesis to create a prediction task to determine if the clusters can predict overall student score. Decision trees for example would provide educators with an easily interpretable indicator of which learning styles are associated with high and low performance to allow for early remedial action to assist struggling students.
- Research in this thesis could be extended beyond the beginners Grok course group by using the same methodologies on the different courses such as the intermediate or advanced NCSS Grok courses. Course comparison of student behaviour can highlight how students perform on varying levels of programming question difficulty and if the same students expressed different learning styles between the courses.
- This thesis can serve as a foundation for future investigation into how to provide personalised feedback to students based on their learning style. The clusters observed in this thesis can be investigated in terms of the types of programming mistakes students make in

each cluster, which can inform the type of feedback these students would most benefit from. Therefore, high risk students would not only be identified, but also served with semi-personalised feedback throughout the course. Semi-personalised feedback could be integrated into Grok as a new feature for future courses depending on the results of this proposed future work.

Appendix A – Anonymised Australasian Computing Education (ACE) 2022 Conference submission

Exploring student's learning style and procrastination behaviours in an online programming course

FirstName Surname
Department Name
University Name
City, State, Country
email

FirstName Surname
Department Name
University Name
City, State, Country
email

FirstName Surname
Department Name
University Name
City, State, Country
email

ABSTRACT

Through analysing student's learning behaviour, teachers can better identify at-risk students and take remedial action to improve the student's online learning experience and course performance. In this paper, we investigate different types of student behaviour through mining data from a Python programming course. Two main research questions are explored; the first objective is concerned with determining trends in student learning behaviour and how these behaviours impact on the student's mark. The student behaviour is analysed through two lenses: the Felder-Silverman learning style model, and through procrastination based features. The second objective seeks to explore how student submission behaviour changes over the duration of the course. This is explored using Markov chains to determine how students move between clusters throughout their weeks of learning.

CCS CONCEPTS

•Educational Data Mining •Clustering student behavior • Markov chains

KEYWORDS

Student learning behavioural classification, assessment, feedback, automatic grading system, student score, course mark.

1 INTRODUCTION

Programming is becoming an increasingly important skill in today's technological society and online courses can provide students with programming course content as well as automatic feedback on programming submissions. Whilst these systems allow for efficient grading, they are not always as effective at providing tailored feedback to students. Additionally, teachers can struggle to provide effective feedback to students due to the high number of enrolments in online courses and the often time-consuming analysis needed to manually assess each student's program and behaviour. Therefore, there is an opportunity to use data driven methods to provide insights into student's learning behaviour and performance in online courses to better personalize feedback to students.

We examine the behaviour of Australian high-school school students participating nationwide in an online Australian programming course, run over a 5-week period in 2019. This courses is offered at different difficulty levels and provide content to learn Python skills as well as testcases for automatic grading of problem submissions. Students undertaking the course do so either as part of their high school curriculum or optionally for their own learning.

There is often a high drop-off rate of course completion from students after the one-week mark. This presents an opportunity to better understand student learning behaviour during the course to increase course completion and student retention. The two main objectives of this study are as follows:

1. To investigate similarities in student behaviours, in particular in terms of learning style and procrastination, and how these behaviours influence the students' overall course score.
2. To investigate how student behaviour changes over time and how behavioural shifts can impact drop off rates.

Overall, this study aims to gain greater insight into student behaviour and what drives course engagement to provide more personalised submissions feedback to students.

2 PREVIOUS WORK

Educational data mining in a range of contexts has sought to identify groups of similar students to determine their behaviour and overall course performance. Various methods have been used – clustering-based [3-5, 8-10], classification [6-7], questionaries and statistical methods [1].

Graf et al. [1] drew from the Felder-Silverman Learning Style Model (FSLSM) which characterizes students into different learning styles: active/reflective, sensing/intuitive, visual/verbal and sequential/global learners. These learning styles determine how a student prefers to learn, for example sequential learners prefer to start and finish tasks in order, whilst intuitive learners prefer high level thinking and will complete multiple tasks simultaneously. FSLSM was utilised in [1] to gain insights into the learner's navigational behaviour in the context of an online object orientated modelling course. This study utilised the Index of Learner Styles (ILS) questionnaire developed by Felder and Soloman in 1997 [2] to determine a student's learning styles. It proved a number of assumptions on the types of learners and how they use an online learning system. For example, detail orientated sensing learners used the forum more than intuitive learners and did so primarily to clarify details of the assessments. The also navigated more to quiz pages since they generally prefer to be well prepared when applying theory to new practical problems. Intuitive learners were shown to go from one challenge or test to another without necessarily completing the previous ones as they prefer not to focus on details.

Conversely, Hooshyar et al. [3] used procrastination features to develop a novel algorithm called PPP to predict student procrastination behaviour to an accuracy of 96%. Procrastination was defined by spare time and inactive time. Spare time is the time from when a student submits an assignment until the assignment is due, and inactive time is the time between when an assignment is open and a student views it for the first time. Using K-means, clusters were categorised into non-procrastinating students, potential procrastinators and procrastinators. Statistical correlation showed there was a positive correlation between spare time and assignment scores and a negative correlation between assignment score and inactive time.

Similarly, Cerezo et al. [4] used procrastination features and effort features to define student learning behaviour in online courses. They considered student behaviour in terms of procrastination levels using six variables: time spent on quizzes, time spent viewing theoretical content, time spent on forums, number of words written on forums, number of other relevant actions and number of days students took to hand in tasks. Using K-means, their results showed students who spent more time on the quizzes performed better than students who did not and that final exam score was not related to time spent on viewing theoretical content or time spent on lecture slides.

Other studies have observed that student behaviour is fluid and can change over time. McBroom et al. [5] captured student's movement between clusters by first clustering students based on behavioural attributes, and then re-clustering based on how many weeks a student's submission belonged to each of the 5 original clusters. Their results concluded the student's behaviour did not change drastically throughout the semester and behavioural clustering and prediction could occur as early as in Week 3 of a 13 week-long semester with accurate prediction of a student's final course mark. Additionally, students who performed in the lowest range of students were consistently in the low performing clusters throughout, but the highest performing students moved to on average, 4.2 clusters out of 6 clusters throughout the semester.

Other studies, e.g. [6], investigated whether at-risk students in programming courses can be detected early in the semester by using information from different sources, e.g. an autograding system, discussion board and assessment marks. They built a decision tree that was able to achieve high accuracy in predicting the exam mark from information available in the middle of the semester. In [7], the same data sources were used to define the characteristics of the high, average and low performing students and predict exam performance.

In this paper, we investigate student behaviour in computer programming courses. We utilize a data-driven approach – clustering and Markov chains, to identify distinct characteristics and trends of student behaviour, and how they influence the overall course score and drop off rate.

3 DATA

All data sourced for this research is from the Beginner stream of the 2019 programming course. This stream was selected due to having a high number of learners and also because the beginners stream engages a wider range of students of different abilities, from first time programmers to accomplished programming students.

The course consists of 5 weeks of content and graded problems. The content (slides), introduces students to the Python concepts which are tested in the graded problems. The slides also can contain programming questions for students to practice their programming skills, where they are allowed unlimited attempts without being penalised. There are 8 scored problems for each of the 5 weeks of the course and these problems are marked by auto grading testcases. The highest possible score in the course is 400 with a maximum of 10 possible marks awarded for each correct question. Marks are deduced after every 5 incorrect submissions a student makes. These deductions are capped at 5 marks, so even a student with a very high number of submission attempts will achieve 5 marks once they pass all the test cases for a question. If a student does not attempt a question or is unable to pass all testcases they receive a 0 score for that problem. A graded submission is different to a terminal run, where a terminal run allows a student to infinitely run their program for a problem without feedback from the testcases and without penalty.

The collated data includes both problem submission data as well as navigational data on how a student utilises the website.

3.1 Feature selection

Feature selection in this study draws from previous studies which utilises FSLSM and procrastination features, and applies them to an online programming course context. Each feature vector represents a student for one week. Therefore, a student has a total of 5 vectors, one for each of the 5 weeks of the course.

The FSLSM learning styles are outlined in Table 1 and the FSLSM features selected are outlined in Table 2. FSLSM features in Table 1 and 2 are based off of Graf et al.'s [1] results and insights on each type of learner. The FSLSM theory describes eight different learning styles (active/reflective, sensing/intuitive, visual/verbal and sequential/global learners) and how each learning style indicates a student's navigational behaviour and interaction with the course website. One feature which was most indicative of each type of learner was selected for this study and adjusted to fit the type of the course. All FSLSM learning groups will be used except for the visual and verbal learning styles due to the difficulty in measuring visual or verbal preference in an online course without student self assessment. It should be noted that while these features have been informed by the FSLSM learning theory, it is not expected that the clusters will strictly fit these behaviour groupings.

Table 1. FSLSM learning behaviour categories and hypothesis on how the categories apply to the course

FSLSM behaviour learning category	Category description	Hypothesis
Active learners	Learn by trying things in a hands-on way. Prefers working collaboratively.	Frequent programming submissions.
Reflective learners	Cerebral learners, learns through thinking and reflecting.	Takes longer on problems to reflect on feedback.
Sensing learners	Prefers concrete examples, practical learners, high attention to detail.	Completes majority of slide questions.
Intuitive learners	Prefers abstract concepts and are innovative. Attempts challenges without completing previous challenges.	Works on several problems at once.
Sequential learners	Learns in linear steps and follows course content sequentially.	High number of sequential slide views.
Global learners	Learns in large leaps and prefers a high degree of freedom in the learning process. Prefers learning from conclusions and summaries.	High number of course menu views to and provide a high-level overview of the course.

Table 2. FSLSM behaviour features

FSLSM learning style	Feature	Description
Active	Number of terminal runs	Number of times a student runs their problem program attempts.
Reflective	Average time spent on problem	Average time spent on problems for the week from first to last submission.
Sensing	Percentage of completed slides	Percentage of content slides with all questions answered correctly.
Intuitive	Problems attempted simultaneously	Represented as a binary with 0 for no problems attempted simultaneously and 1 for problems attempted simultaneously.
Sequential	Percentage of sequential slide views	Slide views which are immediately before or after the current slide viewed, out of all of the slides viewed by a student for the week.
Global	Number of course menu views	Number of times a student viewed the course menu overview.

The selected procrastination features are outlined in Table 3. Following Cerezo's et al. [4] research, both time related features and effort related features have been included to measure procrastination.

Table 3. Procrastination features

Feature	Description
Inactive time	Amount of time between when an assignment is released to a student's first problem submission. Calculated in seconds.
Time spent on content	Amount of time from first slide view to last slide view. Calculated in seconds.
Average time spent on problems	Average time between first and last problem run for each problem in a week. Calculated in seconds.
Percentage of completed submissions	Percentage of submissions which passed all testcases.
Number of autosaves	Number of automatic autosaves. Autosaves are generated after periods of inactivity on the website when programming problems.
Number of terminal runs	Number of times a student runs their programming problem.
Percentage of slides completed	Percentage of slide problems completed. Slide problems are not scored and are questions relating to the concepts taught each week.
Percentage of slides viewed	Percentage of slides viewed by the student for the week.

There is some overlap between the features with terminal run sum, percentage of slides completed and average time spent on problems in both FSLSM and procrastination categories.

4 ANALYSING STUDENT BEHAVIOUR

Similarities in student behaviour were analysed in this study in two ways: using the FSLSM behavioural features, and using the procrastination features. Two separate K-means clusterings were used, one on the FSLSM features and another clustering on the procrastination features. During data cleansing, any vectors with NaN features were removed, resulting in 15733 vectors for the FSLSM clustering, and 19724 vectors for the procrastination clustering. All features were normalised to the range 0 to 1.

4.1 FSLSM behavioural clustering

FSLSM behavioural clustering yielded k=4 different clusters, with the distribution of vectors shown in Table 4 and the centroids of these clusters shown in Table 5. Outliers were removed from features to improve results and increase meaningfulness of the centroids. Despite outliers being removed, for time features like average time spent on problems, there is still a very large range of values from 0 to 518400 seconds which is equivalent to 6 days. Majority of the time values are on the lower end of this range and has resulted in the centroids for time values being small after normalisation.

Table 4. Distribution of vectors in FSLSM clusters

	Number of vectors	Percentage of vectors
Cluster 0	3485	22.15%
Cluster 1	2759	17.54%
Cluster 2	8598	54.64%
Cluster 3	891	5.67%
Total	15733	

Table 5. FSLSM behaviour cluster centroids

	Cluster				
	Full data	0	1	2	3
Terminal runs	0.122	0.087	0.199	0.108	0.154
Slides completed	0.671	0.246	0.849	0.831	0.241
Consecutive slides viewed	0.745	0.726	0.727	0.765	0.679
Menu views	0.114	0.074	0.196	0.101	0.134
Average time spent on problems	0.087	0.083	0.182	0.049	0.171
Simultaneous problems	0	0	1	0	1

From these clusters we can gain the following insights.

Cluster 0: This cluster is characterized by the lowest number of terminal runs which represents the number of terminal runs made before making a scored submission to a problem. Students in this cluster also had a low number of slides completed averaging around 24% of slides completed despite viewing at least 72% of the slides on average. The high number of consecutive slide viewership would be consistent with a Sequential learning style. However, the low number of slides completed may indicate that students primarily read the content slides instead of interacting with the slide questions, or students may have skimmed through the slides without reading them.

Cluster 1: This cluster has the highest number of slides with completed questions with 84.9% of slide questions completed successfully. They had the highest number of course menu views which is a hallmark indicator of Global learners. These learners also worked on more than one problem at once which can indicate Intuitive learning style. These students also work the longest on problems on average, which may be due to them working on more than one problem at once.

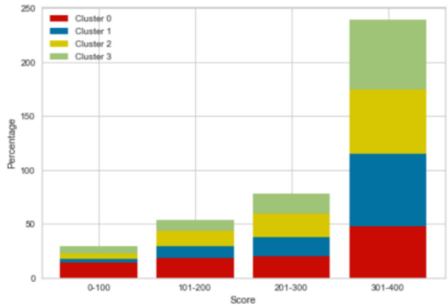
Cluster 2: Similarly to Cluster 1, this cluster also has a high number of completed slides with 83.1% of slides completed. Whilst this cluster is similar to Cluster 1, the main difference is evident in Cluster 2 having the lowest amount of time spent on problems and that on average students complete one problem at a time. This cluster also had the highest number of vectors and consisted of 54.64% of the overall data.

Cluster 3: Cluster 3 has a high number of terminal runs and time spent on problems, indicating they have a high engagement with the problems. However, they were not engaged with the content information and had the lowest slide completion and slide views. They also had high course menu views which could indicate skipping to problems and not engaging with the slide content.

4.1.1 FSLSM behaviour scores

From the clusters, we can determine how student learning behaviour impacts a student's score. Figure 1 shows the score distribution across the clusters as a percentage. It is evident that each cluster has a similar distribution of students of all performance types including low, medium and high performing students. Therefore, the FSLSM features would not be a strong predictive measure of score, as also indicated by the medium to low correlation coefficient of 0.349. However, this score distribution does reflect the literature on the FSLSM learning theory which does not claim that any one group of students will necessarily perform better than others.

Figure 1. Percentage of FSLSM behaviour clusters in each score range



4.2 Procrastination clustering

The second clustering was completed using the procrastination features using K-Means. Using all procrastination features produced a low Silhouette Score of 0.326 and Davies Bouldin score of 1.039 where a Silhouette Score of 1 is ideal and indicates little overlap of the clusters, and low Davies Bouldin score indicates unique clusters. To improve the cluster results, the features were split into two, where one cluster uses time features and the other uses effort related features.

4.2.1 Time procrastination clustering

Time features consist of inactive time, time spent on content, average time spent on problems and percentage of completed submissions. K-means was used to cluster the data and 4 clusters were created based on the elbow method. Table 6 shows the distribution of the vectors in each cluster.

Table 6. Distribution of vectors in time procrastination clusters

	Number of vectors	Percentage of vectors
Cluster 0	3170	16.07%
Cluster 1	8182	41.48%
Cluster 2	2749	13.94%
Cluster 3	5623	28.51%
Total	19724	

Table 7. Time procrastination cluster centroids

		Cluster			
	Full data	0	1	2	3
Inactive time	0.043	0.077	0.044	0.024	0.031
Time spent on content	0.264	0.166	0.088	0.701	0.357
Percentage of completed submissions	0.876	0.349	0.980	0.968	0.976
Average time spent on problems	0.054	0.012	0.003	0.006	0.005

The following clusters were identified from Table 7.

Cluster 0 – High procrastination: High inactive time which meant students waited longer to start problems after they were released compared to the other clusters. They also had the lowest number of completed submissions with only 34.9% of submissions completed. The high amount of time spent on problems can indicate that these students struggled with the content which is why they had a low submission completion rate. Therefore, we can consider this group to have high procrastination tendencies.

Cluster 1 – Low Procrastination: This cluster shows little evidence of procrastination because it has the lowest amount of time spent on content and problems whilst achieving a 98% problem completion rate. This could indicate this group had prior programming experience which allowed them to complete submissions quickly without heavy reliance on content or problem feedback.

Cluster 2 – Low procrastination: Started problems shortly after the release time and had the highest amount of time spent on content. This cluster also had a high amount of completed submissions with 96.8% of submissions completed.

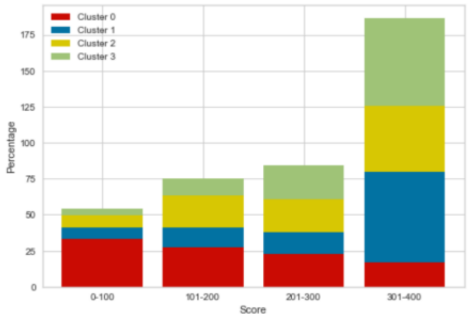
Cluster 3 – Moderate procrastination: High number of completed submissions and a medium amount of time spent on content. This cluster also had a moderate level of inactive time compared to the other clusters.

4.2.1.1 Procrastination time scores

From the scores in Figure 2, we can see that all clusters except for Cluster 0 show a similar distribution of marks in the low, medium and high range. This score distribution supports the cluster results since Cluster 0 (high procrastination) has the lowest problem completion rate, with only 16% of scores achieved in the 301-400 range. Interestingly, whilst Cluster 2 (low procrastination) and Cluster 3 (moderate procrastination) had similar centroids for most features, Cluster 2 had more students in the low to mid score range, with 54% of vectors between 0-300 marks, whilst Cluster 3 had 39% of students in this same score range. The predominant difference between these two cluster's centroids is students in Cluster 3 spent less time on content and slides. This can indicate that Cluster 3, like Cluster 1 had students with previous programming knowledge who didn't have to rely on the slide content as heavily as other students.

Additionally, Cluster 3 (moderate procrastination) had 39% of students in the 0-300 range which is similar to Cluster 1 (high procrastination) which had 36% of students in the low to mid-tier score range. Therefore, despite Cluster 3 having higher procrastination tendencies they were still able to achieve similar results to students with low procrastination tendencies.

Figure 2. Percentage of procrastination time clusters in each score range



4.2.2 Effort procrastination clustering

Effort related features consist of number of autosaves, number of terminal runs, percentage of completed submissions, percentage of completed slides, percentage of slides viewed. Whilst procrastination features are usually time related, Cerenzo's et al [4] procrastination clusters also drew from other effort related features to measure the level of student course engagement that occurred during the time students were on the platform. The elbow method indicated k=5 clusters to be used for this set of K-means clustering, with the distribution of the clusters shown in Table 8, and centroids in Table 9.

Table 8. Distribution of vectors in effort procrastination clusters

	Number of vectors	Percentage of vectors
Cluster 0	12135	61.52%
Cluster 1	2596	13.16%
Cluster 2	1808	9.17%
Cluster 3	1443	7.32%
Cluster 4	1742	8.83%
Total	19724	

Table 9. Effort procrastination cluster centroids

	Cluster					
	Full data	0	1	2	3	4
Number autosaves	0.089	0.998	0.089	0.039	0.081	0.067
Terminal runs	0.057	0.064	0.058	0.025	0.050	0.048
Comple-ted sub-missions	0.876	0.983	0.973	0.276	0.461	0.947
Slides completed	0.647	0.839	0.267	0.292	0.727	0.176
Slides viewed	0.842	0.962	0.864	0.412	0.823	0.429

The effort procrastination clusters are described below.

Cluster 0 – High problem effort, high content effort: Cluster 0 has the highest amount of engagement on all measures, with 97% of submissions completed, 83% of slides completed and 96% of slides viewed. They also had a high number of terminal runs and autosaves which suggests they spent a significant amount of effort on the problem component of the course.

Cluster 1 – High problem effort, medium content effort: This cluster is characterised by a low number of slides completed despite having a high number of slides viewed with 86% of slides viewed. This suggests students in this cluster may skim through the content without engaging with the practice questions.

Cluster 2 – Low problem effort, low content effort: Cluster 2 has very low activity on all features is seen in this cluster. These students likely didn't engage with the course and had the lowest number of completed submissions.

Cluster 3 – Low problem effort, high content effort: This group of students had a low number of completed submissions, on average only completing 46% of the problems. However, they did have high slide viewership and the highest slide completion rate with over 72% of slides completed.

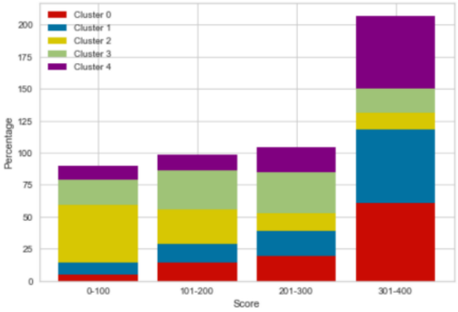
Cluster 4 – High problem effort, low content effort: These students had a low number of autosaves and terminal runs but had a high completion rate of submissions with an average of 94% of problems completed. Whilst a low number of autosaves and terminal runs would indicate a lack of engagement with the course problems, the high number of completed problems for this groups indicates that this cluster of students may have been quicker and more efficient at solving problems, perhaps because of prior programming knowledge obtained before completing the course. This cluster can also be characterized by a very low amount of slide content interactions, with only 17% of slide questions completed and 42% of slides viewed.

4.2.2.1 Procrastination time scores

Regarding the scores obtained by students in each cluster in Figure 3, Cluster 2 (low problem effort, low content effort) and Cluster 3 (low problem effort, high content effort) have the lowest number of students obtaining a high mark in the 301-400 range. This is consistent with the centroids of these two clusters which shows low problem completion for both groups.

Three clusters have high achieving students: Cluster 0 (high problem effort, high content effort), Cluster 1 (high problem effort, medium content effort), and Cluster 4 (high problem effort, low content effort). Cluster 0 has the strongest results with 61% of students achieving a score of 301-400, and Cluster 1 and 4 having 56% of students in the same score range. Cluster 0 has strong problem and content effort metrics which likely contributed to their higher marks.

Figure 3. Percentage of procrastination effort clusters in each score range



Cluster 1 and 4 distinguish themselves by their interaction with the slides, with Cluster 1 showing high levels of slide viewership and low levels of slide completion, whilst Cluster 4 has low metrics in both slide measures. Despite both clusters having 56% of students achieving high marks, Cluster 4 shows 11% of students obtaining a mark of 0-100, while Cluster 1 has 8% of students in this score range. This can indicate that whilst some students in Cluster 4 could have previous knowledge allowing them to score highly without engagement with the course content, this cluster may also include some students who did not engage, and therefore contributing to the lower marks in this cluster.

4.3 Cluster overlap

Whilst undertaking separate clustering can highlight the different groups of students regarding their FSLSM learning style, or levels of procrastination, investigating the overlap between the clusters can provide a greater understanding of how students undertook the course wholistically.

The behavioural clusters and procrastination time clusters were selected for clustering overlap analysis due to them best fitting the two theories from previous work and because they have the same number of clusters (k=4) allowing for more seamless analysis.

Table 10 shows the largest overlap of the clusters is between behaviour Cluster 2 and procrastinate Cluster 1 (low procrastination), Cluster 2 (low procrastination) and Cluster 3 (moderate procrastination). This may be because behaviour Cluster 2 consists of 55.67% of all the vectors for that clustering. All of the procrastination Clusters 1, 2 and 3 have a high number of submissions completed but varying number of time spent on problems and content but high scores in the upper ranges of 301-400. This can suggest that students of behaviour Cluster 2 are also low to mid-range procrastinators.

Table 10. Number of overlapping vectors in FSLSM behaviour clusters and procrastination time clusters

		FSLSM behaviour clusters			
		Cluster 0	Cluster 1	Cluster 2	Cluster 3
Procrastination time clusters	Cluster 0	1040	154	747	83
	Cluster 1	1306	1134	3459	438
	Cluster 2	371	419	1527	110
	Cluster 3	761	1052	3865	260

Procrastination Cluster 0 (high procrastination) had the lowest scores, started problems late, had a high amount of time spent on problems and a low completion rate. This cluster is mostly overlapped with behaviour Cluster 0. Behaviour Cluster 0 had low number of terminal runs as well as low slide completion. This could indicate that whilst these students spent a long time on problems, most of this time was not spent on terminal runs and therefore students may not have been running their Python program prior to submission.

By considering the overlap of both sets of clusters, we can gain greater insight into the behaviours of students of different procrastination levels. This can help teachers identify not only which behaviours can lead to procrastination but also how to best remediate this procrastination behaviour.

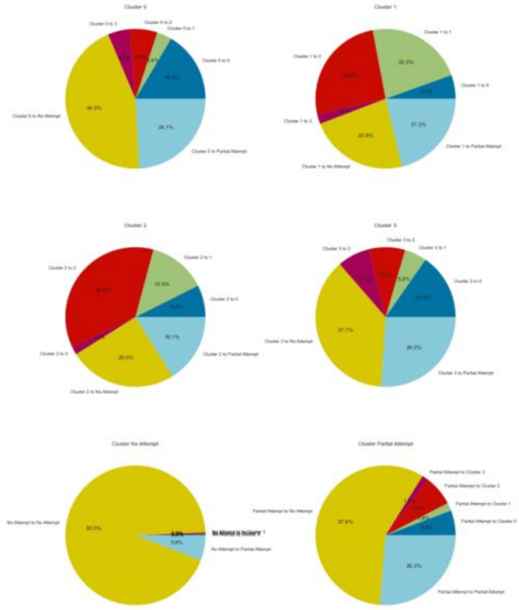
5 ANALYSING CHANGES IN STUDENT BEHAVIOUR OVER TIME

Student learning is not static and a student's behaviour can change over time throughout the course duration. Markov chains were used to analyse how students moved between clusters each week. An aggregated version of the Markov chain in Figure 4, 5 and 6, shows which clusters the students from each Week 1 cluster transitioned to in any of the proceeding weeks. In addition to the clusters generated in Section 4, two other clusters were added to the Markov chain. One cluster is No Attempt which means a student did not have any activity on the website for that week. In other words, the vector for that week would be all NaN. The second cluster is Partial Attempt which means that students were engaged in some, but not all features measured. This can help indicate to teachers when a student drops out of the course and when to best engage students to prevent drop-out rates.

5.1 FSLSM behaviour Markov chain

Figure 4 shows the Markov chain for the FSLSM behaviour clusters. The Cluster 0 pie chart in Figure 4 shows that 44.5% of vectors in Cluster 0 in Week 1 transitioned to No Attempt in any of the proceeding weeks. This is the highest percentage transition to No Attempt of any of the FSLSM clusters. This indicates that Cluster 0 students are of high risk of dropout in the course. Additionally, the Cluster 1 and 2 pie charts in Figure 4 shows high movement between Cluster 1 and 2. These two clusters are similar, except Cluster 1 spends more time on problems than Cluster 2 and Cluster 1 works on more than one problem at once. The relatively high number of transitions between Cluster 1 and 2 suggests that students can switch between behaviour types as the content increases in difficulty.

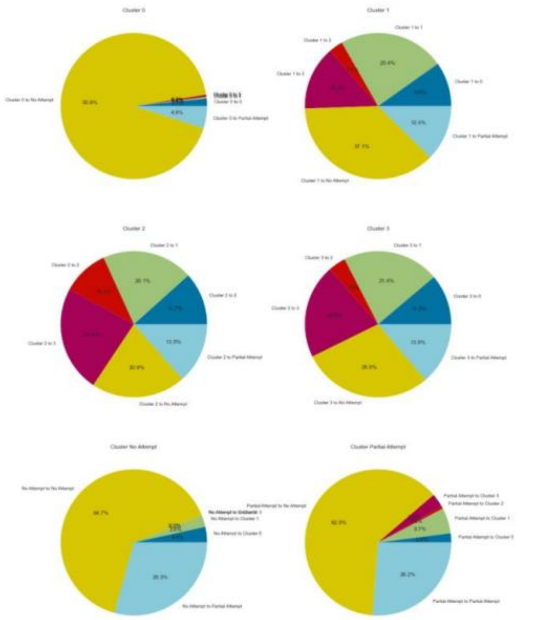
Figure 4. FSLSM behavioural Markov chain. Percentage of transitions from Week 1 clusters to clusters in proceeding weeks



5.2 Procrastination time Markov chain

The procrastination time Markov chain from Figure 5, shows a particularly high drop off rate for Cluster 0 (high procrastination) with 92% of students who were in Cluster 0 in Week 1 moving to the No Attempt cluster in the proceeding weeks. Cluster 2 (low procrastination) and Cluster 3 (moderate procrastination) show a relatively smaller number of drop off rates but a high number of transitions to other clusters. Both Cluster 2 and 3 are similar, where students start questions early shortly after they are released, but with Cluster 2 spending almost double the average time on content than Cluster 3. The interchangeableness of these clusters means students may be spending more or less time on content depending on their skill level and the difficulty of the content for the week.

Figure 5. Procrastination time Markov chain. Percentage of transitions from Week 1 clusters to clusters in proceeding weeks



5.3 Procrastination effort Markov chain

In Figure 6, Cluster 2 (low problem effort, low content effort) and Cluster 3 (low problem effort, high content effort) were the students of the highest risk of dropout with 95% and 81% of students who began in this cluster moving to No Attempt in previous weeks respectively. This is higher than the students who began in the No Attempt group in Week 1. Whilst this could be expected for Cluster 2, which had low activity in all effort features, Cluster 3 did have 72% of slides viewed on average, which would normally indicate high engagement with the content. In this example, the Markov chain highlighted at-risk students who may have gone unnoticed due to high measures on some features.

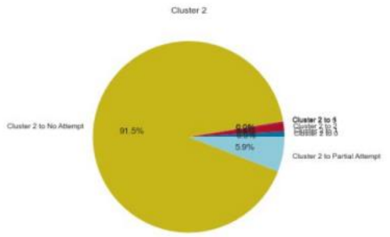
Figure 6. Procrastination effort Markov chain. Percentage transition of clusters from Week 1 to clusters in proceeding weeks

Overall, each Markov chain shows a pattern that once a student has made no attempt for the week, they are unlikely to re-engage with the course.

5.4 Changes in student learning week by week

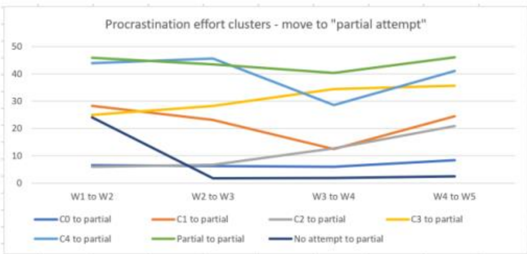
The Markov chains can also provide more detailed insights when calculated between each consecutive week. An example of a consecutive week Markov chain for the procrastination effort features is shown in Figure 7. Previously in Figure 6 it was identified that Cluster 2 (low problem effort, low content effort) and Cluster 3 (low problem effort, high content effort) are the highest risk students of disengaging with the course. By using a consecutive week Markov chain we can determine not only which students are most at-risk, but when they are most at-risk of disengaging with the course. For example, when looking at procrastination effort transitions between Week 1 and Week 2 in Figure 7, we can see that Cluster 2 had a 91.5% drop off rate after the first week. Therefore, these students would require intervention as early as during the first week of the course to prevent their course engagement levels from dropping off.

Figure 7. Procrastination effort Markov chain from Week 1 to Week 2



The consecutive Markov chains can also indicate the difficulty that students experienced with the course content each week. In Figure 8 between Week 3 to Week 4, students had a drop in partial attempts which indicates that students found Week 4 easier and were able to complete more of the content and questions. It should be noted that the students from Cluster 3 and Cluster 2 who saw an increase in partial attempts in this week also had a drop in No Attempt in the same week. Therefore we can assume these students who were previously not attempting questions from these clusters were able to attempt questions in Week 4 due to the easier content.

Figure 8. Procrastination effort clusters to Partial Attempt clusters between each consecutive week of the course.



6 CONCLUSION

This paper investigates different ways to compare the similarities and differences in student's learning approaches, using both the FSLSM learning style and procrastination features. For the FSLSM features, 4 clusters were generated which reflected students of different learning styles where all clusters had a similar distribution of low, medium and high scores. Hence, students with various learning styles were able to achieve good performance. Conversely, the procrastination features were divided into two groups for clustering, the first group relating to time features and the second group relating to effort features. The time clusters identified one high, one moderate and two low levels of procrastination groups.

Clustering of effort related procrastination features formulated 5 different clusters relating to student's effort towards problems and slide content. These procrastination clusters had more distinct distribution of scores between each clusters with the low procrastination and the 'high problem effort, high content effort' clusters representing the students with the highest scores in the 301-400 range. Analysing the overlap of the vectors of different clusters also proves useful for gaining more wholistic insight into the type of procrastination tendencies students in each behavioural clusters and visa versa.

The Markov chain results allowed for insight into the student drop off rates in each set of clusters and can identify at-risk students throughout the course duration. Our results showed that students with low procrastination effort measures were at most risk of having no attempt in the course in previous weeks. Using Markov chains to compare the student's cluster transition between consecutive weeks can highlight when a group of students is most at-risk of dropping out of the course. It was shown that the low effort student's participation dropped off as early as after Week 1. Therefore, the Markov chains can inform teachers when to provide assistance to at-risk students to prevent drop off rates as it was noted that in all clusters once a student disengages in the course for a week it is unlikely they will reengage in any proceeding weeks.

Overall, our method allows teachers to identify groups of poorly performing students based on their learning behaviours and can highlight when these students are most at-risk of becoming disengaged from the course. These insights can help teachers with remedial action and submission feedback to assist struggling students and to target the students who require assistance to improve student learning outcomes and performance in the course.

7 REFERENCES

[1] Graf, S., Lui, T. C., Kinshuk., 2010. Analysis of learner's navigational behaviour and their learning styles in an online course. *Journal of Computer Assisted Learning 26,* 2, 116-131

[2] Felder, R. M., Soloman, B. A. (1997). Index of learning styles questionnaire. Available at: http://www.engr.ncsu.edu/ learningstyles/ilsweb.html. Accessed: 2021-04-03

[3] Cerezo, R., Esteban, M., Sanchez-Santillan, Nunez, J. C., 2017. Procrastinating behaviour in computer-based learning environments to predict performance: A case study in Moodle. *Frontiers in Psychology*, *8,* 1403.

[4] Hooshyar, D., Pedaste, M., Yang, Y., 2019. Mining educational data to predict student's performance through procrastination Behaviour. *Entropy*, 12, 1-24. Available at: https://www.mdpi.com/1099-4300/22/1/12

[5] McBroom, J., Jeffries, B., Koprinska, I., Yacef, K, 2016. Mining behaviours of students in autograding submission system logs. In *Proceedings of the International Conference on Educational Data Mining,* 159-166.

[6] Koprinska, I., Stretton, J., and Yacef., K., 2015. Students at risk: detection and remediation. In Proceedings of the *International Conference on Educational Data Mining*, 512-515.

[7] Koprinska, I., Stretton, J., and Yacef., K., 2015. Predicting student performance from multiple data sources. In Proceedings of the *International Conference on Artificial Intelligence in Education*, LNCS 9112, 678-681.

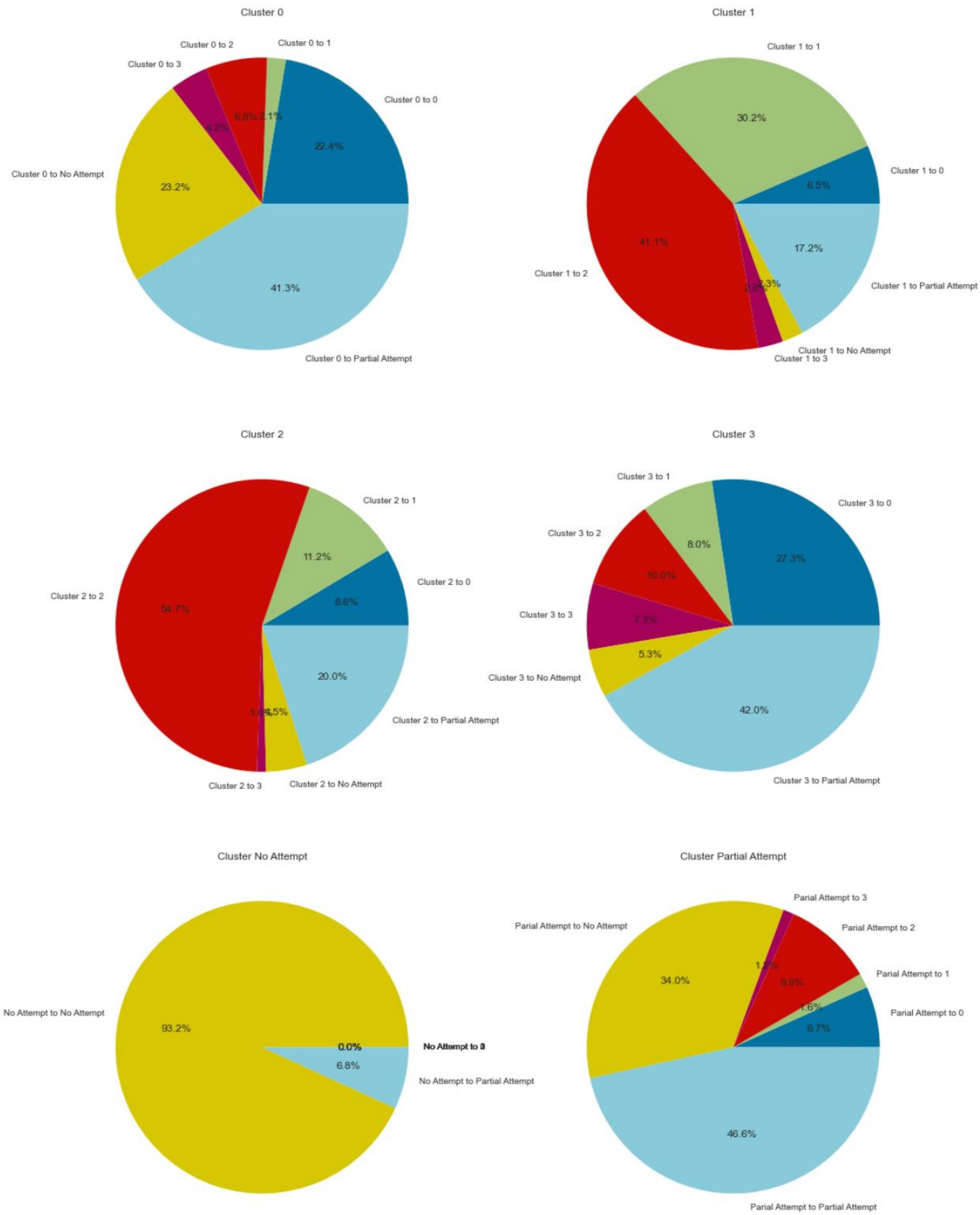
[8] Glassman, E., 2016. Clustering and visualizing solution variation in massive programming classes. Ph.D. Dissertation. Massachusetts Institute of Technology.

[9] Gulwani, S., Radiček, I., and Zuleger, F., 2018. Automated Clustering and Program Repair for Introductory Programming Assignments. ACM SIGPLAN Notices 53(4), 465-480.

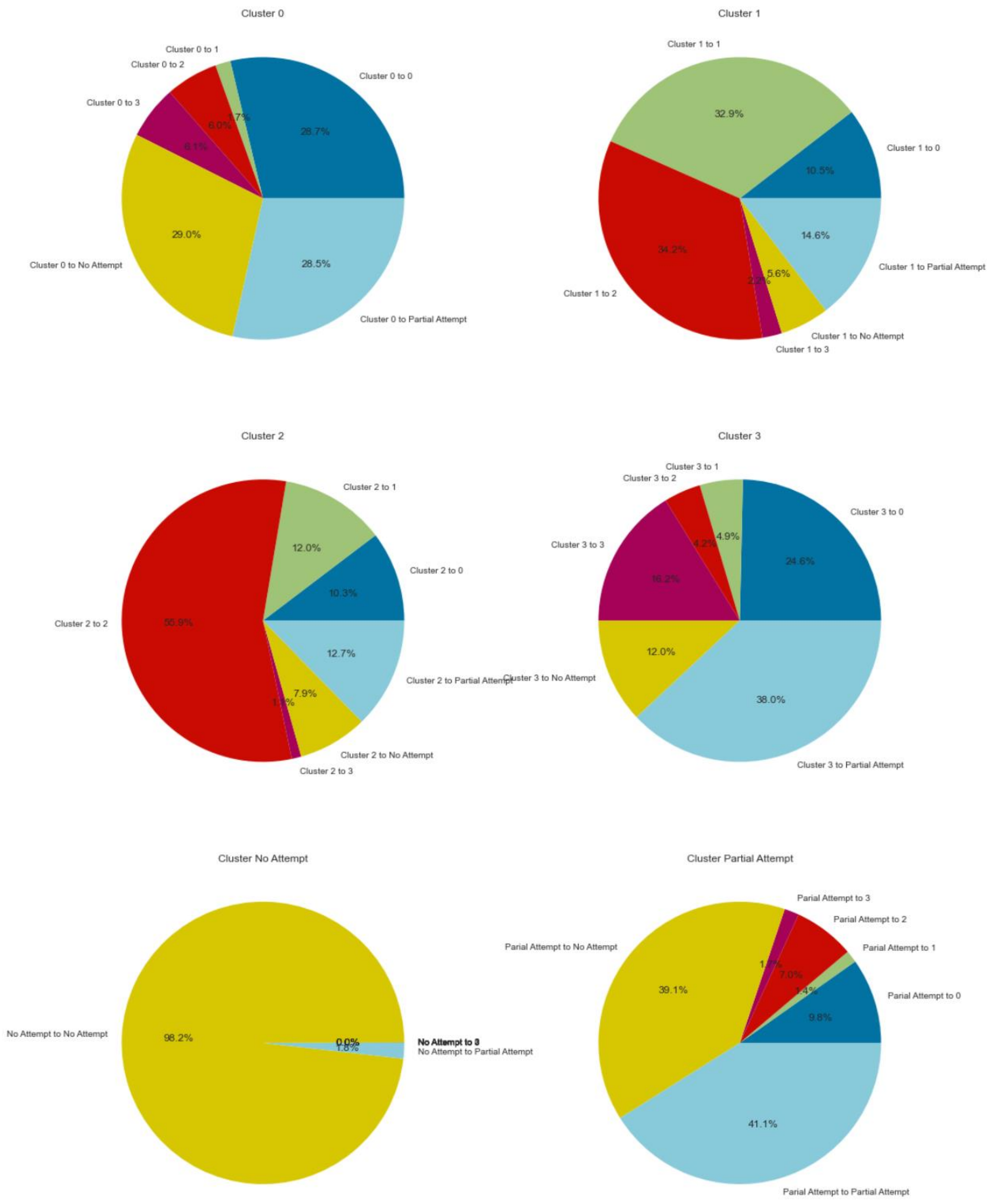
[10] McBroom, J., Yacef, K. and Koprinska I., 2020. DETECT: A Hierarchical Clustering Algorithm for Behavioural Trends in Temporal Educational Data. In Proceedings of the *International Conference on Artificial Intelligence in Education*, 374-385.

Appendix B – FSLSM behavioural Markov chain pie charts for consecutive weeks

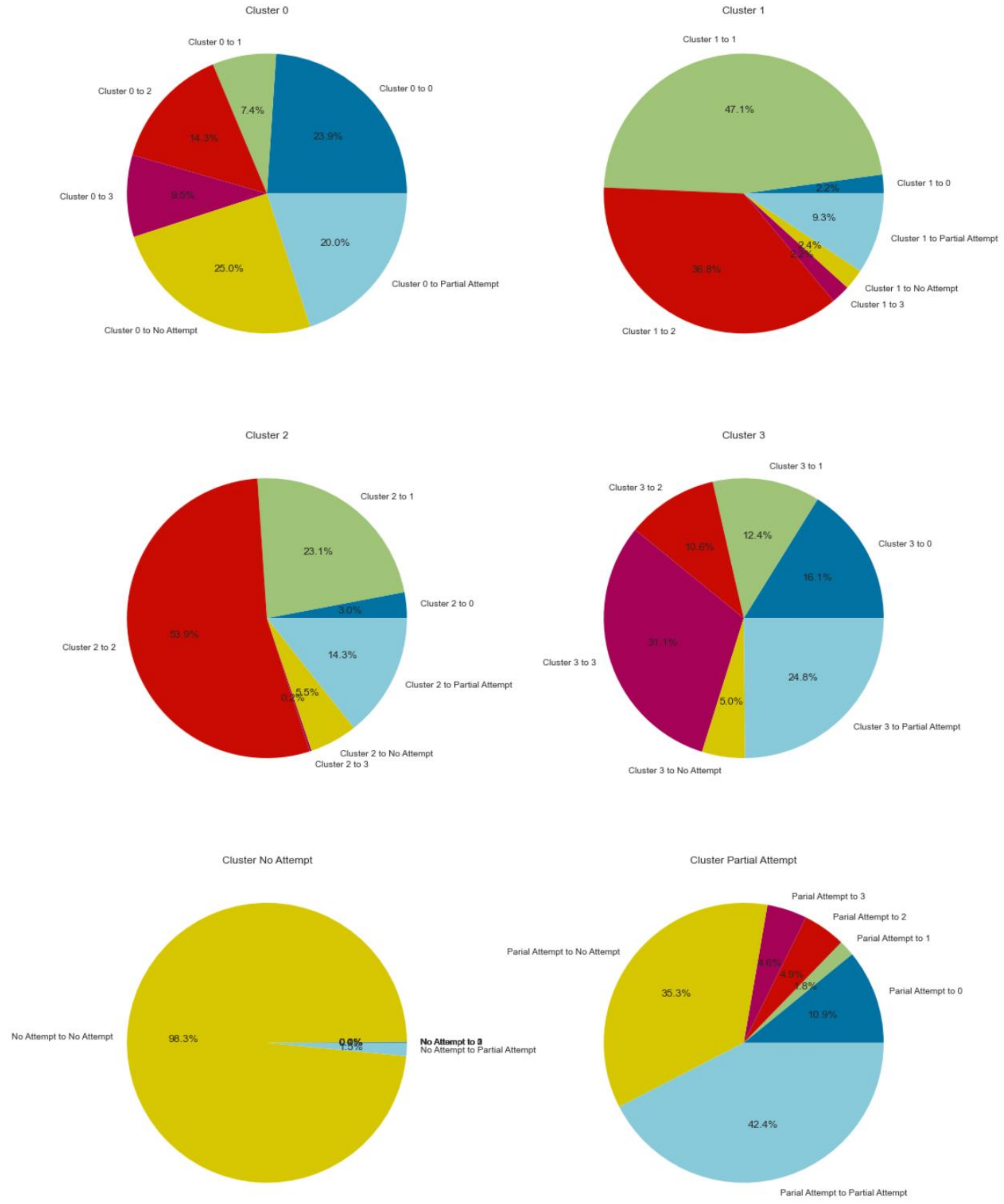
B1 FSLSM behavioural Markov chain pie chart Week 1 to Week 2



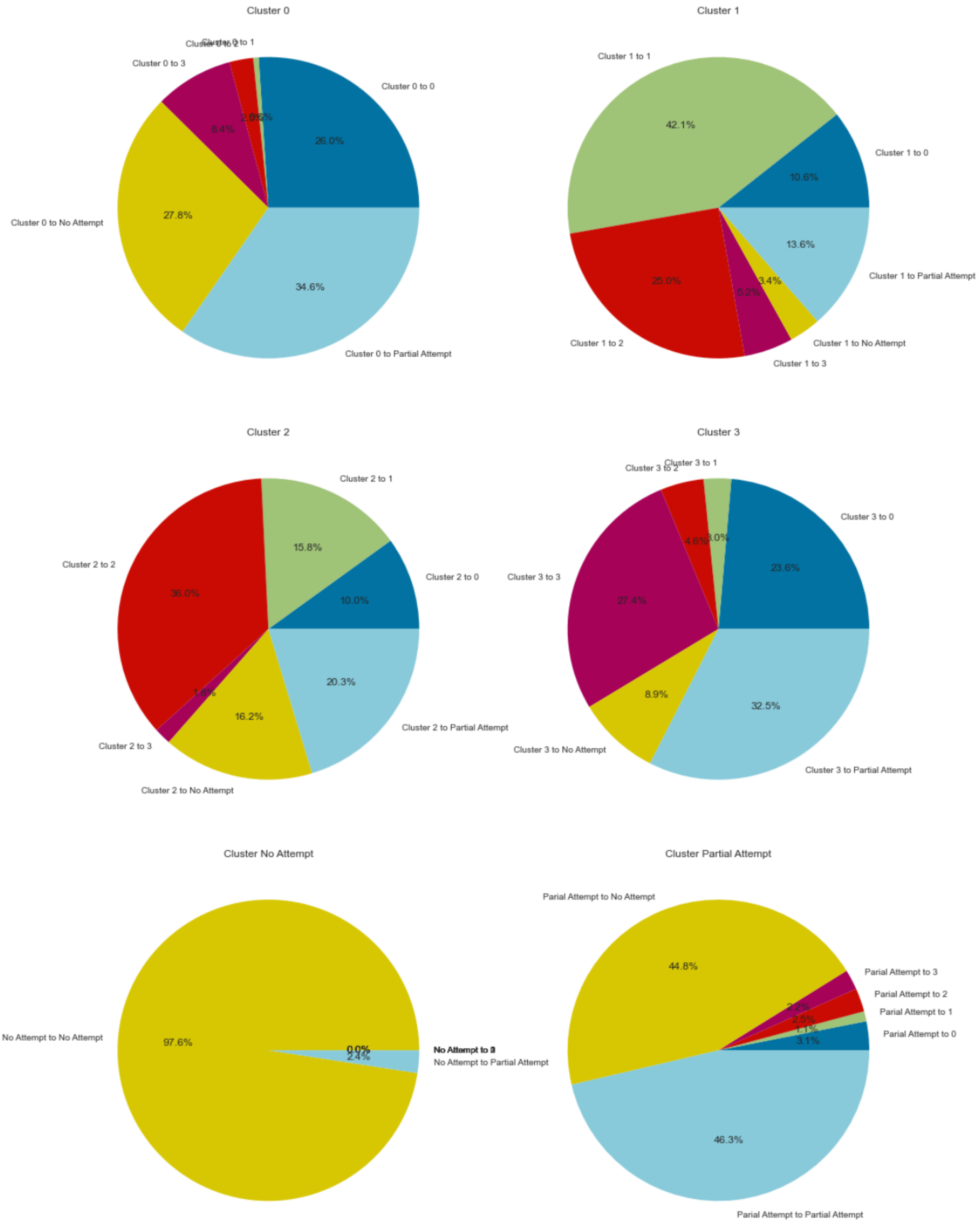
B2 FSLSM behavioural Markov chain pie chart Week 2 to Week 3



B3 FSLSM behavioural Markov chain pie chart Week 3 to Week 4

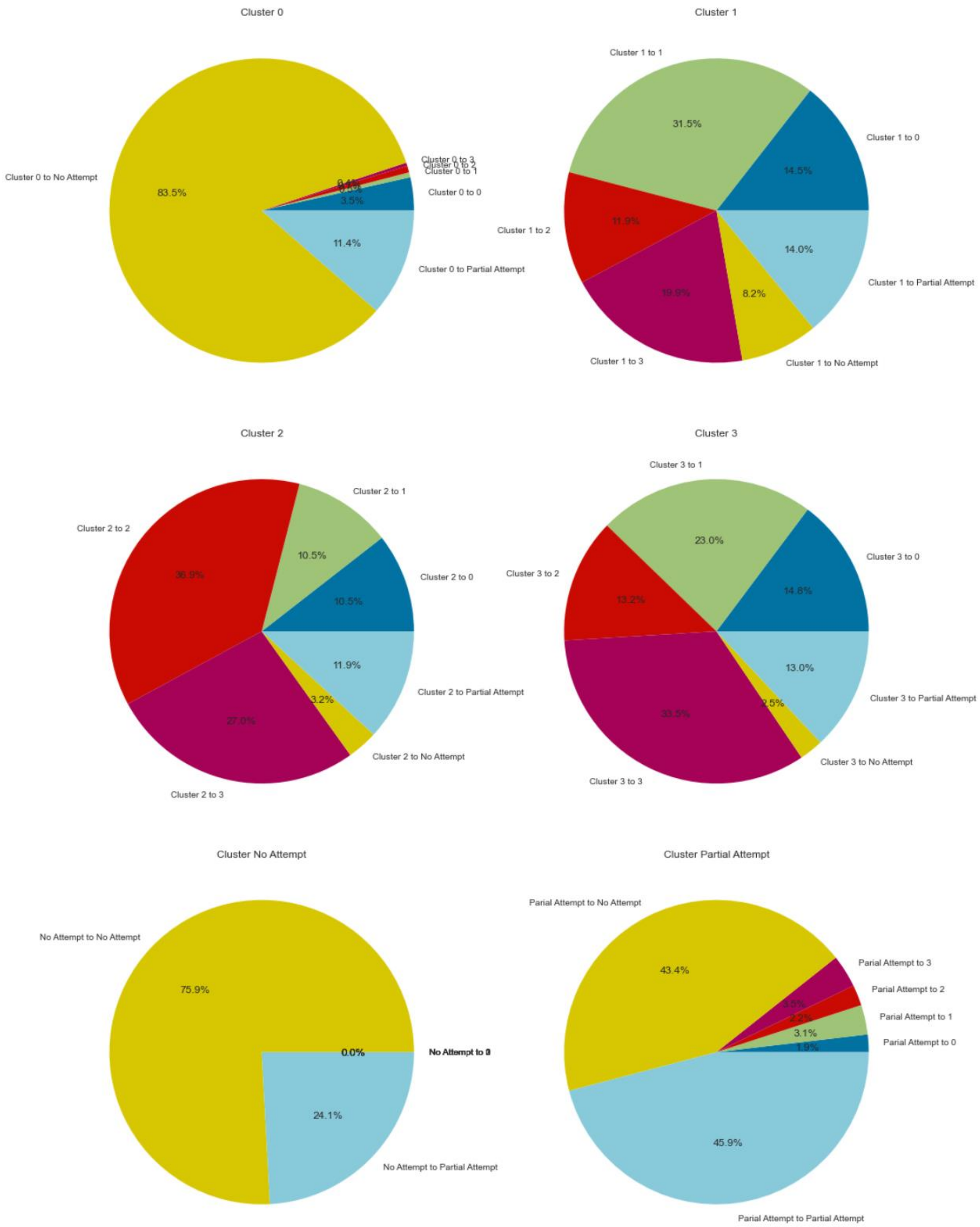


B4 FSLSM behavioural Markov chain pie chart Week 4 to Week 5

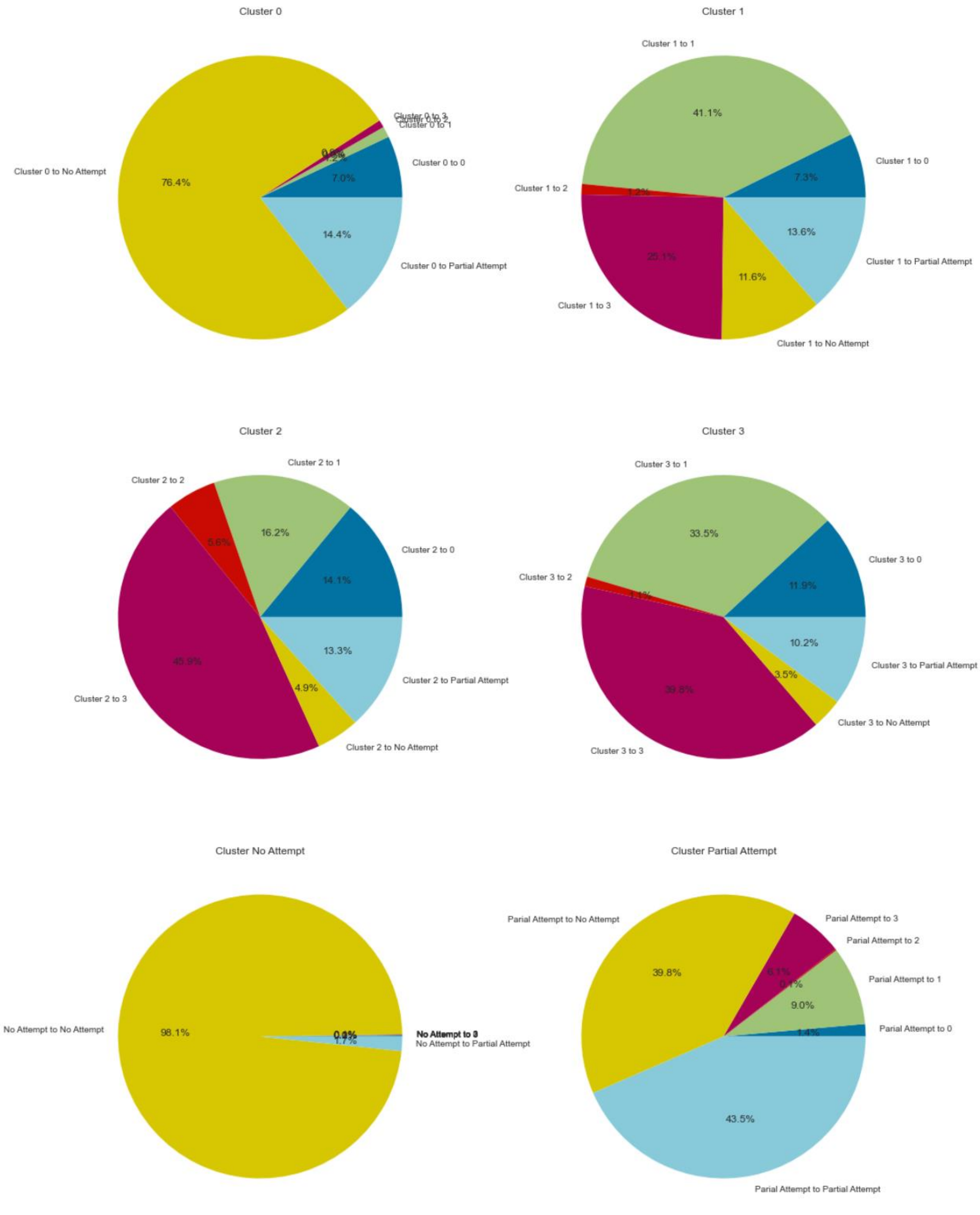


Appendix C – Procrastination time Markov chain pie charts for consecutive weeks

C1 Procrastination time Markov chain pie chart Week 1 to Week 2

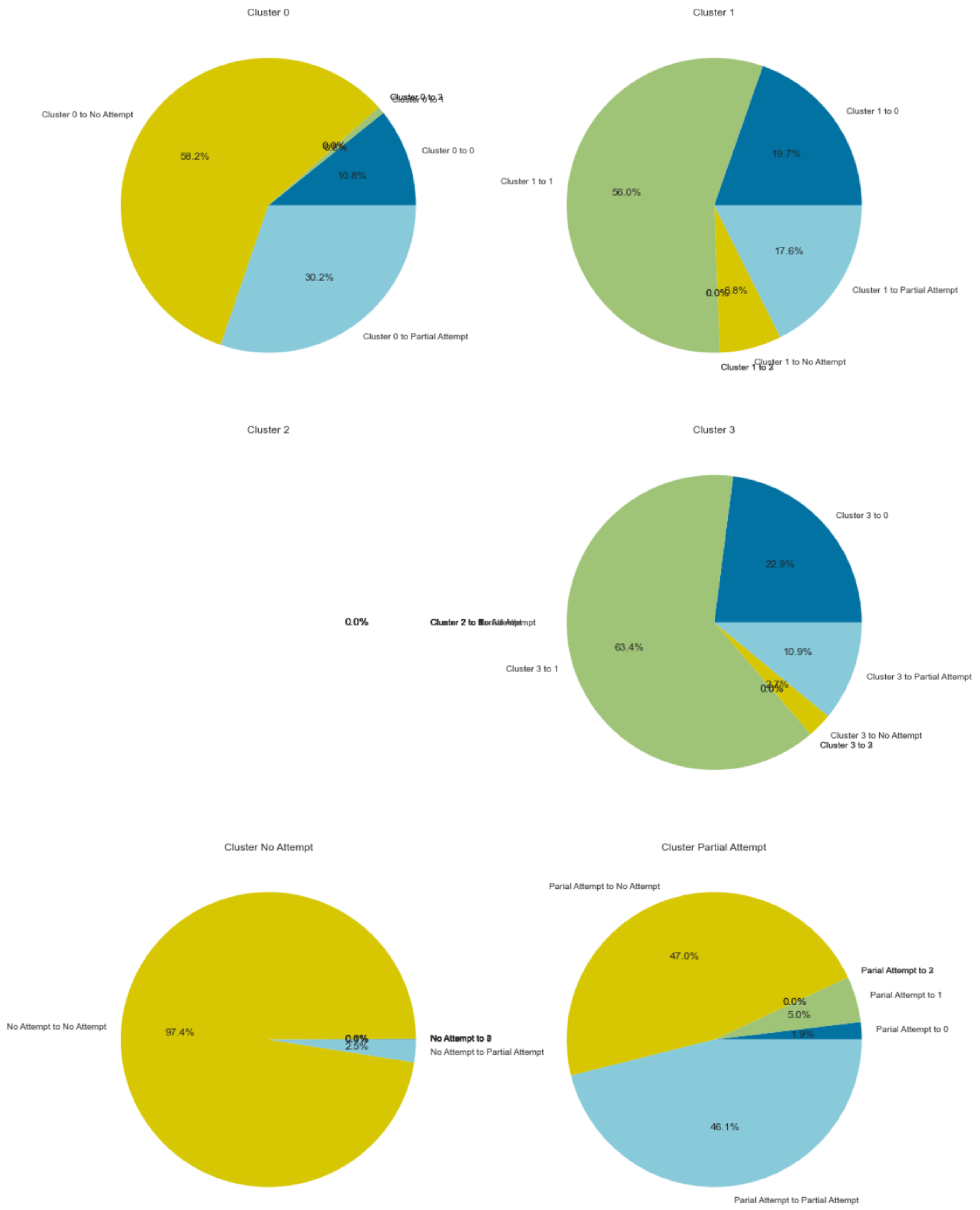


C2 Procrastination time Markov chain pie chart Week 2 to Week 3



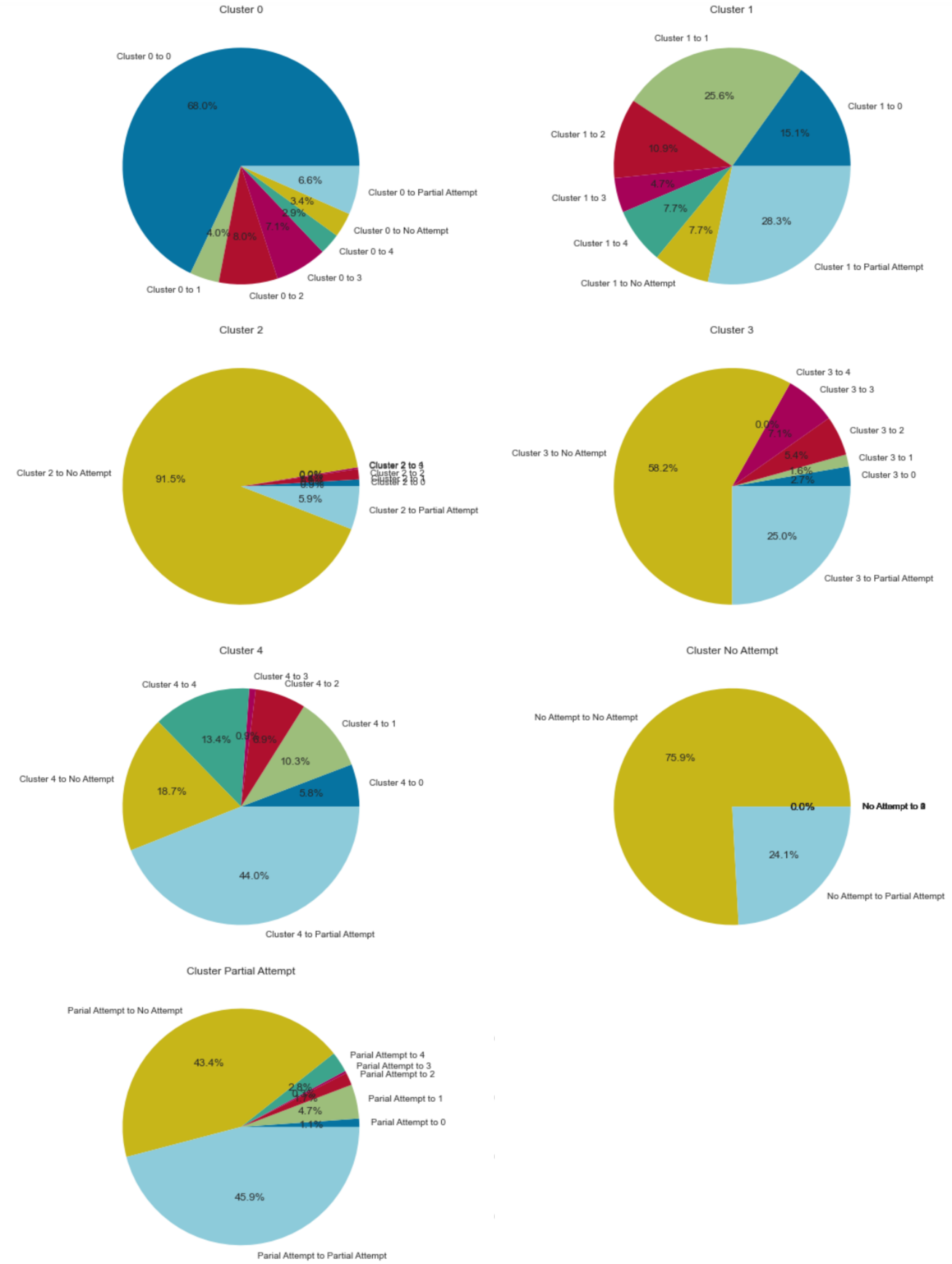
C3 Procrastination time Markov chain pie chart Week 3 to Week 4

C4 Procrastination time Markov chain pie chart Week 4 to Week 5

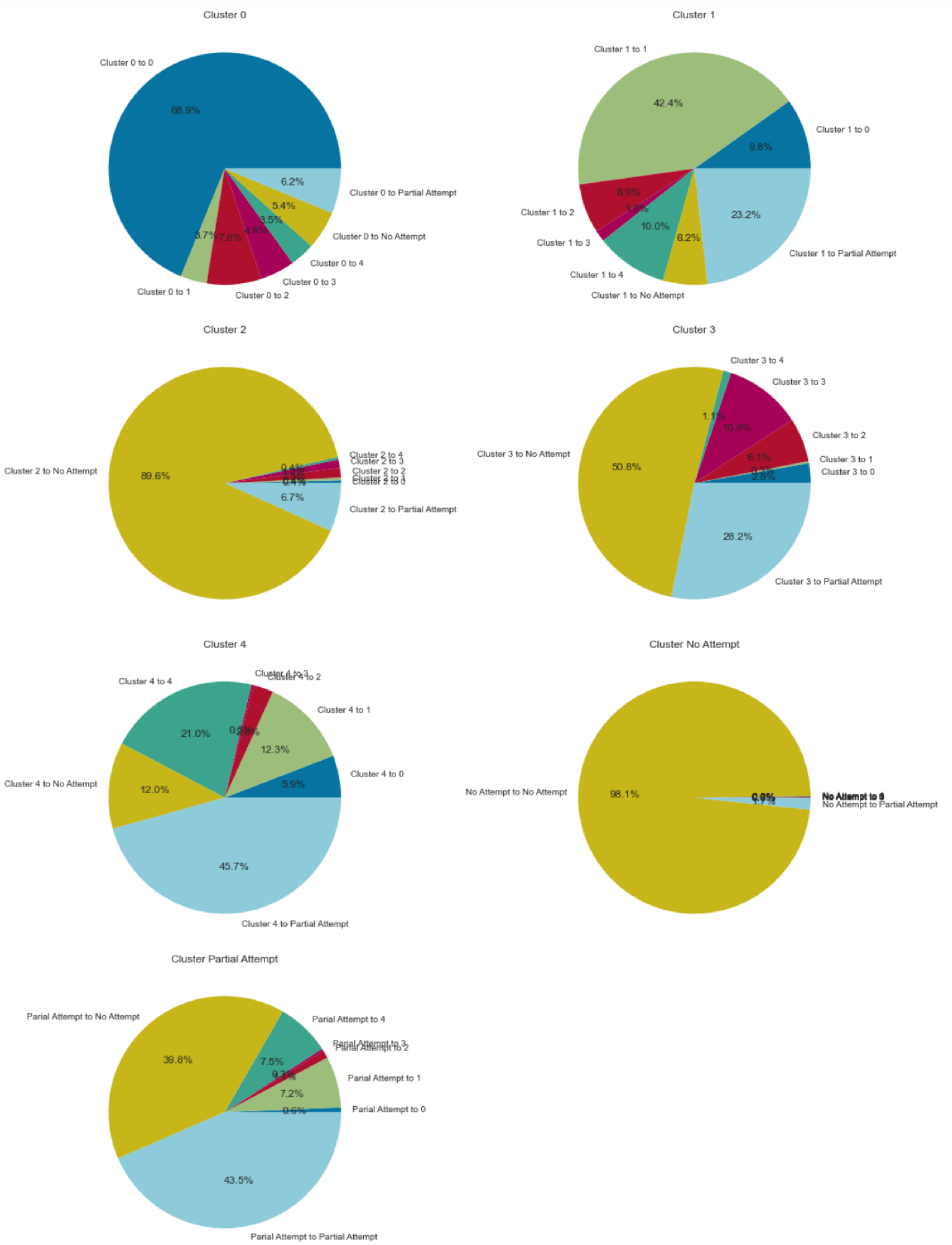


Appendix D – Procrastination effort Markov chain pie charts for consecutive weeks

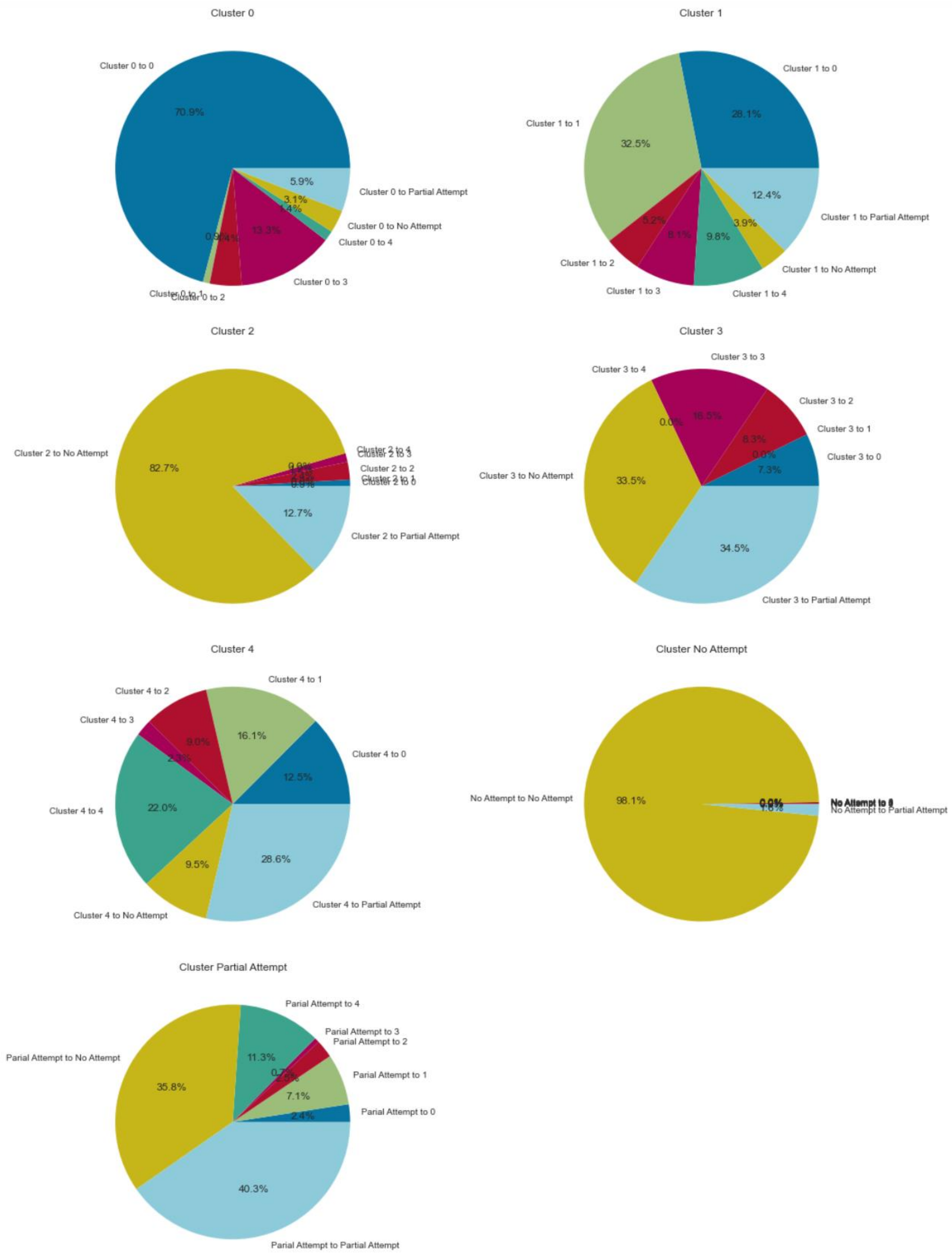
D1 Procrastination effort Markov chain pie chart Week 1 to Week 2



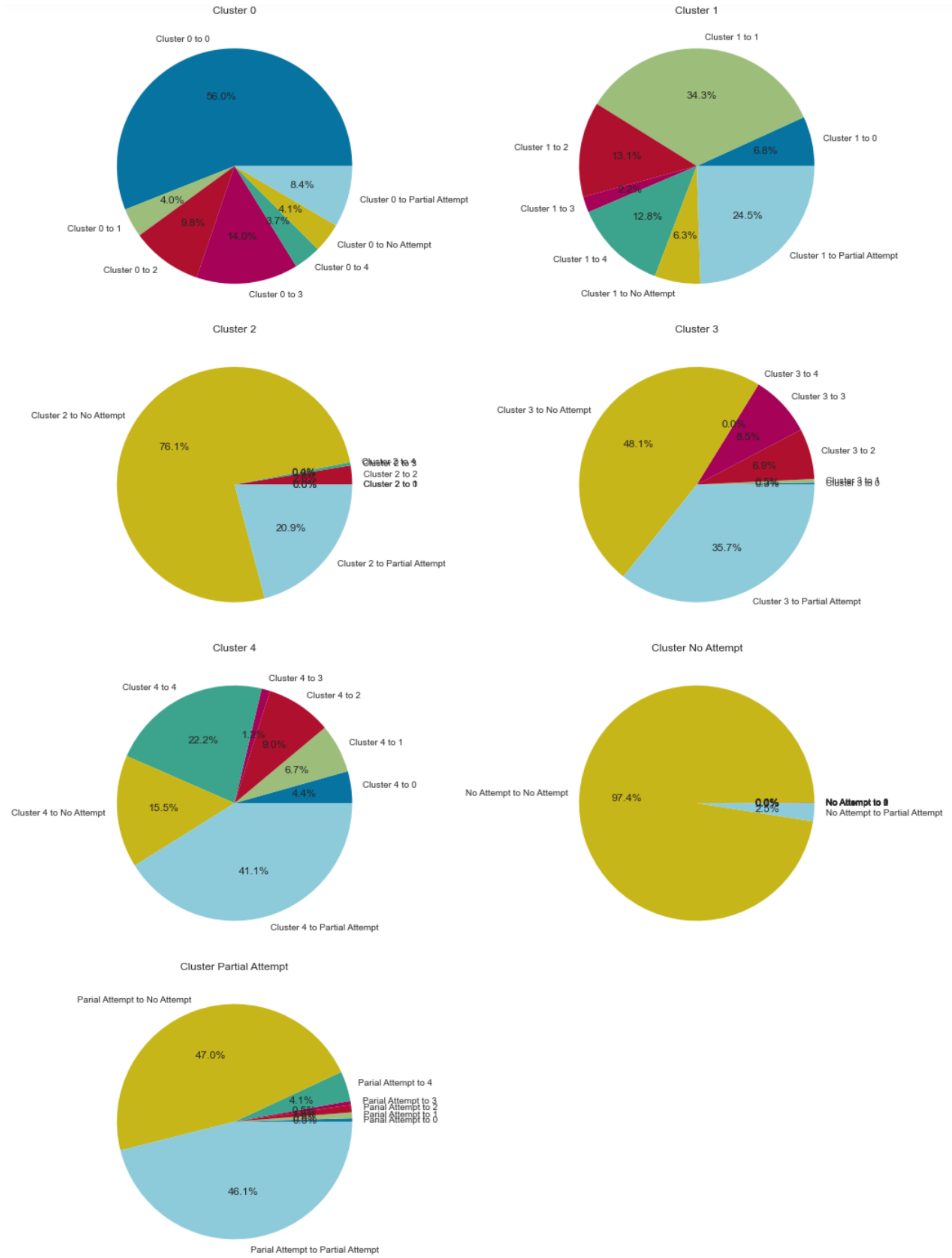
D2 Procrastination effort Markov chain pie chart Week 2 to Week 3



D3 Procrastination effort Markov chain pie chart Week 3 to Week 4



D4 Procrastination effort Markov chain pie chart Week 4 to Week 5



References

[1] A. Burkov, “Chapter 5: Basic Practice,” in *The Hundred-Page Machine Learning Book,* 2019.

[2] P. Tan, M. Steinbach, A. Karpatne, and V. Kumar, “Cluster Analysis: Basic Concepts and Algorithms,” in *Introduction to Data Mining,* Pearson, 2nd ed, 2018.

[3] F. Suthar, “A Study on Educational Data Mining”. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 7, no. 2, pp. 1089-1096, Feb. 2019, doi: 10.22214/ijraset.2019.2172.

[4] B. Ugalde, V. Radhakrishnan, “A Research Travelogue Towards Educational Data Mining,” *International Journal of Computer Applications,* vol. 179, no. 42, pp. 39-45, May. 2018, doi: 10.5120/ijca2018917005.

[5] V. Kumar, E. Stracke, “Examiner’s reports on theses: Feedback or assessment?,” *Journal of English for Academic Purposes,* vol. 10, pp. 211-222, 2011, doi: 10.1016/j.jeap.2011.06.001.

[6] J. Biggs, “Individual Differences in Study Processes and the Quality of Learning Outcomes,” *Higher Education,* vol. 8, pp. 381-394, July. 1970, doi: 10.1007/BF01680526.

[7] L. Li, C. Tsai, “Accessing Online Learning Material: Quantitative Behaviour Patterns and their Effects on Motivation and Learning Performance,” *Computers & Education*, vol. 144, pp. 286-297, July. 2017, doi: 10.1016/j.compedu.2017.07.007.

[8] M. Hartnett, A. George, J. Dron, “Examining Motivation in Online Distance Learner Environments: Complex, Multifaceted, and Situation-Dependent,” *International Review of Research in Open and Distributed Learning,* vol. 12, no. 6, pp. 20-38, Oct. 2011, doi: 10.19173/irrodl.v12i6.1030.

[9] R. M. Felder, L. K. Silverman, “Learning and Teaching Styles in Engineering Education,” *Engineering Education*, vol. 78, no. 7, pp. 674-681, 1988.

[10] S. Graf, T-C. Lui, Kinshuk, “Analysis of learner’s navigational behaviour and their learning styles in an online course,” *Journal of Computer Assisted Learning*, vol. 26, no. 2, pp. 116-131, January. 2010, doi: 10.1111/j.1365-2729.2009.00336.x.

[11] R. Cerezo, M. Esteban, M. Sanchez-Santillan, J. C, “Procrastinating Behaviour in Computer-Based Learning Environments to Predict Performance: A Case Study in Moodle,” *Frontiers in Psychology,* vol. 8, no. 1403, Aug. 2017, doi: 10.3389/fpsyg.2017.01403.

[12] D. Hooshyar, M. Pedaste, Y. Yang, “Mining Educational Data to Predict Student’s Performance through Procrastination Behaviour,” *Entropy,* vol. 22, no.12, pp. 1-24, Dec. 2019, doi: 10.3390/e22010012

[13] J. McBroom, B. Jeffries, I. Koprinska, K. Yacef, “Mining behaviours of students in autograding submission system logs,” *International Conference on Educational Data Mining,* pp. 159-166, Jun. 2016.

[14] S. M. S. Chow, “Generating Data-driven Feedback for Computer Programming Students,” The University of Sydney, 2016.

[15] I. Koprinska, J. Stretton, K. Yacef, “Students at Risk: Detection and Remediation. University of Sydney,” School of Information Technologies, 2015.

[16] J. McBroom, B. Jeffries, I. Koprinska, K. Yacef, “Exploring and Following Student’s Strategies when Completing their Weekly Tasks,” *EDM,* University of Sydney, School of Information Technologies, 2016.

[17] J. McBroom, I. Koprinska, K. Yacef, “Understanding Gender Differences to Improve Equity in Computer Programming Education,” *ACE ’20: Twenty-Second Australasian Computer Education Conference,* pp. 185-194, Feb. 2020, doi: 10.1145/3373165.3373186.

[18] M. Macedo, H. V. Siqueira, E.M.N Figueiredo, A. M. A. Maciel, “Clustering students based on grammatical errors for on-line education,” *Journal of Brazilian Society on Computational Intelligence (SBIC),* vol. 16, no. 1, pp. 26-40, June. 2018, doi: 10.21528/LNLM-vol16-no1-art2.

[19] J. McBroom, K. Yacef, I. Koprinska, “DETECT: A Hierarchical Clustering Algorithm for Behavioural Trends in Temporal Educational Data,” *International Conference on Artificial Intelligence in Education (AIED),* pp. 374-385, June. 2020, doi: 10.1007/978-3-030-52237-7_30.

[20] D. Perry, R. Samanta, D. Kim, X. Zhang, “SemCluster: Clustering of Programming Assignments based on Quantitative Semantic Features,” *PLDI ’19: 40th ACM SIGPLAN Conference on Programming Language Design and Implementation* (*PLDI)*, pp. 860-873, June. 2019, doi: 10.1145/3314221.3314629.

[21] S. Kaleeswaran, A. Santhiar, A. Kanade, S. Gulwani, “Semi-Supervised Verified Feedback Generation”, *FSE ’16: 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering,* pp. 739-750, Nov. 2016, doi: 10.1145/2950290.2950363.

[22] S. Sharma, P. Agarwal, P. Mor, A. Karkare, “TipsC: Tips and Corrections for Programming MOOCs,” *Artificial Intelligence in Education,* June. 2018, doi: 10.1007/978-3-319-93846-2_60.

[23] J. McBroom, B. Paaben, B. Jeffries, I. Koprinska, K. Yacef, “Progress Networks as a Tool for Analysing Student Programming Difficulties,” *ACE ’21: Australasian Computing Education Conference,* pp. 158-167, February. 2021, doi: 10.1145/3441636.3442366.

[24] R. Cerezo, M. Esteban, M. Sanchez-Santillan, J. C. Nunez. “Student’s LMS Interaction Patterns and their Relationship with Achievement: A Case Study in Higher Education,” *Computers & Education*, vol. 96, pp. 42-54, May. 2016, doi: 10.1016/j.compedu.2016.02.006.

[25] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, “Clustering,” in *Data Mining: Practical Machine Learning Tools and Techniques,* 4th ed, 2011.

[26] A. C. Muller, S. Guido, “Clustering,” in *Introduction to Machine Learning with Python,* O’Reilly Media, Inc. 2016.

[27] Scikit-yb, “Elbow Method,” https://www.scikit-yb.org/en/latest/api/cluster/elbow.html, (accessed Oct. 9, 2021).

[28] M. Gilli, D. Maringer, E. Schumann, “Modelling dependencies”, in *Numerical Methods and Optimization in Finance*, ScienceDirect, 2nd ed, 2019.

[29] J. D. Kelleher, B. M. Namee, A. D’Arcy, “Chapter 10 Beyond Prediction: Unsupervised Learning,” in *Fundamentals of Machine Learning for Predictive Data Analytics*, 2nd ed, pp. 613-616, 2020.

[30] Pandas, “pandas.DataFrame,” https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html, (accessed Aug. 12, 2021).

[31] Scikit-learn, “sklearn.metrics.silhouette_score,” https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html, (accessed Sept. 19, 2021).

[32] Scikit-learn, “sklearn.metrics.davies_boulding_score,” https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html, (accessed Sept. 19, 2021).

[33] P. Bruce, A. Bruce, P. Gedeck, “Correlation” in *Practical Statistics for Data Scientists*, 2nd ed, 2020.