# Using early assessment performance as early warning signs to identify at-risk students in programming courses

Ashok Kumar Veerasamy
*Department of Future Technologies*
*University of Turku*
Turku, Finland
0000-0002-2942-0303

Daryl D'Souza
*School of Science & IT*
*RMIT University*
Melbourne, Australia
daryl.dsouza@rmit.edu.au

Mikko-Ville Apiola
*Department of Future Technologies*
*University of Turku*
Turku, Finland
mikko-ville.apiola@utu.fi

Mikko-Jussi Laakso
*Department of Future Technologies*
*University of Turku*
Turku, Finland
milaak@utu.fi

Tapio Salakoski
*Department of Future Technologies*
*University of Turku*
Turku, Finland
tapio.salakoski@utu.fi

***Abstract*— This full paper presents results of a model developed using early assessment tasks as predictors to identify at-risk students. To date several studies have been conducted to identify and retain at-risk students in computer science courses. However, both researchers and teachers have long sought to understand early warning signs for identifying at-risk students. While coursework-based predictive models have been developed, they need further investigation, due to inconsistencies in a range of identified factors and techniques employed. This paper presents a classification tree analysis (manually created) and a Random forest classification-based predictive model that uses two variables to predict student performance in introductory programming. Visualisation of the decision tree results is employed as early warning signs for instructors to assist students who identified as at-risk. Data for the formative assessment tasks in the first two weeks of the semester was used for model development, validation and testing. The overall prediction accuracy of the model was 60%. The results of this study showed that it is possible to predict 77% of students that need support, as early as Week 3, based on student performance in continuous formative assessment tasks in a 12-week introductory programming course. Moreover, our classification tree analysis revealed that students who secured less than or equal to 25% in formative assessment tasks in the first two weeks are unlikely to attend or indeed fail the final exam. Additionally, the results provide useful insights for early interventions, to prevent attrition and failure and to increase student retention and student success.**

***Keywords—early assessments tasks, at-risk students, parsimonious models***

## I. Introduction

Despite many educational technologies and learning strategies developed by researchers to improve student learning in computer programming, average successful completion rate in introductory programming was 67.7% in 2014 and 71.3% in 2018, and the dropout and failure rates of computing degrees is relatively high in comparison to other university degree programs [1, 2]. Teaching programming remains one of the computing education's main challenges [2]. Several studies have investigated factors that influence student performance in programming in order to improve student programming skills and, in turn, to reduce failure rates [3, 4]. The use of early warning indicators to predict student outcomes has become an emerging focus for schools and higher education institutions, in order to be able to make timely decisions to improve student success [5, 6]. The phrase "at-risk students" is typically used in educational settings to refer to a group of students who struggle with their studies or risk of failing academically, or have a higher probability of dropping out of school. They are usually poor academic achievers, who need academic support from instructors and academic advisors. Increase in at-risk student numbers, course non-completion and student attrition rates, result in university outcomes that concern key stakeholders (students, instructors, course administrators, academic advisors and senior management in institutions). Early warning systems have been developed to use such indicators via educational data mining techniques, to identify students who are at risk and need academic support [7, 8]. Despite these developments, student failure rates continue to be of concern to the computer science education community, and educators are often searching for key factors that may serve as early warning signs or performance indicators to identify at-risk students [9]. The fundamental question is "How might we identify at-risk students early in the semester?" This paper seeks to determine whether early measures of student performance in continuous formative assessments might serve as early warning signs, to identify students that need early intervention, in order to facilitate student retention.

This study was motivated by the need for early identification of students who are at-risk of academic failure. Formative assessment plays a vital role in student learning and achievement [10] and complements summative assessment such as a final exam [11]. The early weeks of formative assessment results provide good opportunities to partially assess student learning outcomes and to identify at-risk students. As such, this study includes performance in ongoing assessment tasks as key indicators for identifying at-risk students. Anecdotal evidence from our faculty suggests that, on average, nearly 30% of students did not attend the end-of-the final exam each year. The need to identify students who need support early in the semester, is therefore crucial for instructors to provide timely interventions, before students reach a point of no return. One way to assist teachers is to provide early warnings visually, presenting patterns of success and failure for prescriptive analysis by instructors. Information visualisation (reporting) is one of the functions of a learning analytics engine, to provide intuitive and easily-understood representations for stakeholders [12]. This study investigates the effectiveness of two prediction methods to identify at-risk students: (i) classification tree analysis (manually created), and (ii) Random forest binary classification based predictive model which is limited to the following independent variables to identify at-risk students:

homework and demo exercises results during the first two weeks of the semester. The impact of other causal factors such as login data, problem solving skills, prior knowledge in programming, and self-esteem is beyond the scope of this study, and will be considered in a future study.

## II. RELATED WORK

We present here related work concerning early warning indicators in higher education, the impact of formative assessment tasks in learning and student achievement, and predictive modelling in computing education.

### *A. Early warning/performance indicators to identify student at-risk of failing*

The term at-risk is widely used in education to describe students who have a high probability of failing academically, and who typically need intervention from instructors to succeed academically. Increase in at-risk students is of concern and identifying at-risk students during the early stages of a course attracts significant interest by educators (instructors, policy makers and institutions) [13]. Studies have been conducted to understand the key precursors of dropping out, to identify students likely to fail a course, in turn to provide early intervention to at-risk students [14, 15]. Students display signs of being at-risk of failure before they actually disengage from their studies [16]. Bruce et al. [17] report irregular attendance, poor course performance, and behaviour problems of students, as early warning signs of students falling off track. Horton [14] listed key factors, including low self-efficacy, taking a course after an extended absence, limited English proficiency, poor prior educational experience, and health and psychological problems, which may compromise academic success or are key indicators of at-risk students at college level. A study on student disengagement concluded that poor attendance, attending classes without completing assignments or note reading, low GPA, failure to submit or late submission or fails in assignments, and not becoming involved with other students in group projects or collaborative learning activities, are directly associated with student disengagement and in turn are key indicators to predict students at-risk of academic failure [18]. However, other studies have shown that the effect of lecture attendance on student learning and performance may vary due to the student demographics and education setup. For example, Horten et al. and Veerasamy et al. [19, 20] reported that student lecture attendance and student performance in formative and summative assessment tasks are negatively correlated. Despite these mixed results, most of these studies confirm the view that assessments are mirrors of what students learn; and instructors therefore use formative assessment tasks as monitoring instruments to discover student learning difficulties [21, 22].

### *B. Formative and summative assessment tasks*

Formative assessment is course assessment used to evaluate and monitor student learning throughout the course. These assessments are mainly used to gauge students' understand of concepts or lack thereof to potentially improve student performance and to tune ongoing teaching and learning strategies. Formative assessment partly determines the final grades of the student. Formative assessment influence student achievement; students who complete formative assessments perform significantly better in the final exam than those who do not [23]. For example, homework assignments are provided during a course of study to enhance student achievement. Fan et al. [24] conducted a 20-year meta-analysis (1986-2015) on homework and students' achievement in mathematics and science and concluded that there was a positive relationship between homework and academic achievement. Similarly, Bas et al. [25] conducted a meta-analysis of homework and academic achievement and found that 64% of all research studies on homework and student achievement revealed that homework has a positive effect on students' academic success, although other studies revealed that homework does not always have, or may have partial impacts on final exam or academic success [26, 22]. Several studies have investigated the importance of formative assessment for student learning and achievement, in science and technology courses, including programming [10, 27, 28]. In addition, studies have included formative assessment as predictor variable to predict student performance [29, 30]. Veerasamy et al. [20] reported that marks attained in homework and demo exercises have a positive correlation with final programming results. Formative assessment tasks such as small-group-discussion based tutorial activities improved student performance in final introductory science exams [31].

Similarly, summative assessment is a form of assessment which is typically conducted one or more points during the semester and at the end of the course duration to evaluate student achievement. Summative assessments are mainly used to identify how much a student learned throughout a course. Examples are mid-semester and final exams, conducted in the middle and at end of the semester to produce end results or a score to compute grades. Summative assessment results are intended to measure a student's level of learning of concepts by a certain point in time. On the other hand, continuous formative assessment results report student ongoing learning and provide a qualitative insight into student learning for teachers to identify students that need support early to provide timely interventions [32]. We can conclude from the literature that scores received in formative assessment might serve as early warning signs, to identify at-risk students. As such, our study uses formative assessment scores as criteria to identify at-risk students for predictive analysis.

### *C. Predictive modelling in computer science education*

Predictive modelling is a process that uses statistical techniques including machine learning algorithms and known results to develop and validate models used for predicting future outcomes. It comes under the category of predictive analytics, which is one of the stages of learning analytics, and widely used in education to predict student performance, enrolment, retention, institutional progress and more. It is based on current and past student and institutional data for stakeholders. There have been several studies conducted on prediction models, which have employed different statistical techniques, including machine learning techniques, as well as different data sources as inputs, to predict students who need support in programming courses [33, 34, 35, 36, 37, 38]. For example, Ahadi et al. [33] explored nine different classification machine learning algorithms using computer science course assignment snapshot data (24 programming assignments conducted in week one) and prior grade average data as input for model development, to predict student final grades in programming courses. However, the collection of such data from students requires significant out-of-class effort in the first week, and such collection practices may not be applied in regular university contexts. In addition, this approach cannot be applied to courses for which grades attained in the prerequisite courses is unavailable for instructors. Porter et al. and Liao et al. [34, 35, 37] used student clicker data in their studies to identify students that needed support early in the term. However, the methodology applied in these studies required the use of clickers and peer-instruction pedagogy and raised concern whether this approach can be applied to courses that do not employ instrumented integrated development environments, and the peer-instruction pedagogy. Moreover, Liao et al. [35] used data imputation with informed guessing in order to not to lose the vast majority of students' data; it raises concern over the validity of the study results as regression-based data imputation might lead to biased parametric estimation. In other work by Liao et al. [38] they used data sources, including clicker data responses, take-home assignment, online quiz grades and final grade from prerequisite courses as input into predictive models to predict student performance in multiple computer science courses. This study used logistic regression claiming that it provided strong prediction accuracy for a binary outcome. The results of this study may be biased as it did not provide any statistical evidence to support the use of logistic regression over other machine learning algorithms. As such, the following points emerged from these studies; (i) the data sources used require significant collaboration effort from students located at different geographical locations, and it may be challenging for instructors to obtain or access data for predictive analysis; (ii) the methodologies used in these studies cannot be applied to courses that use different pedagogy; (iii) it is not yet clear which machine learning algorithms are preferable in this context; (iv) the predictive models did not consider using parsimonious approaches in developing the models; (v) most of the studies have used K-fold cross-validation to evaluate the model's performance; and (vi) several studies examined the influence of formative assessment on students' learning and academic achievement. Hence, we surmise that parsimonious models are simple models that seek to minimise the number of predictor variables, yet likely to perform well on unknown data and formative assessment could be included in a predictive modelling to identify student at-risk of failing.

### *D. Predictive modelling with Random forest classification*

Random forest classification is an ensemble learning method for binary and multiclass classification and regression problems. It is user friendly and computationally less intensive than other ensemble classification methods [39]. Random forest classification has proven that its average prediction accuracy is higher than many other classification algorithms such as K-nearest neighbour, Logistic regression, Naïve Bayes, Support vector machine, Neural networks, and Decision trees [40, 41, 42, 43]. For example, Perez et al. [44] used four data mining algorithms to predict university student drop-out rates and concluded that the best area under the curve in predicting student drop-out rates was achieved by Random forest. However, Bydžovská [45] analysed studies that used different machine learning techniques including Random forest and found that Support vector machine achieved the best performance in predicting student performance over other classification algorithms. On the other hand, other studies have claimed that Random forest classification does not over-fit and is not sensitive to noise or overtraining and can handle high dimensional data [39]. In addition, we explored other machine learning techniques such as Naïve Bayes, C5.0 and Support vector machine for predicting students' performance. Random forest classification provided better overall prediction and at-risk balanced accuracy across our datasets compared to other machine learning models.

In summary, we developed a parsimonious model with explanatory data sources such as student continuous assessment task results, which are easily accessible by instructors for predictive analytics. Our study adopted the Random forest classification algorithm with 10-fold cross-validation, to develop a predictive model to explore the predictive capabilities of selected categories of variables. In addition, our study also promotes a visualisation of student performance results for instructors to identify students at-risk, in order to facilitate timely interventions.

## III. RESEARCH METHODOLOGY

The objective of this study was to identify and visualise at-risk students early who were not qualified to sit the final exam or likely to fail it. The study used a simple methodology using student performance in selected formative assessment tasks, to predict, visualise and to identify at-risk students. The following assessment tasks were used: homework and demo exercises, with the first two weeks of performance scores used as indicators of performance in final programming exam. The identification of at-risk students was conducted via two methods (Fig. 1). Fig. 1 illustrates the research methodology adopted in this research paper to predict students at-risk of failing the final exam.

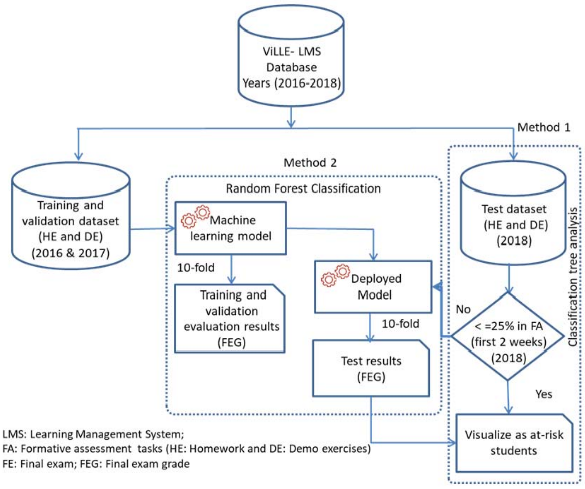
ViLLE- LMS
Database
Years (2016-2018)
Method 1
Method 2
Random Forest Classification
Training and validation dataset (HE and DE) (2016 & 2017)
Machine learning model
10-fold
Training and validation evaluation results (FEG)
Deployed Model
10-fold
Test results (FEG)
Test dataset (HE and DE) (2018)
No
< =25% in FA (first 2 weeks) (2018)
Yes
Classification tree analysis
Visualize as at-risk students
LMS: Learning Management System;
FA: Formative assessment tasks (HE: Homework and DE: Demo exercises)
FE: Final exam; FEG: Final exam grade

Fig. 1. Classification tree analysis and predictive model development, validation, and testing to predict at-risk students

This study visualises poor performing students' early assessment performance based on two methods: (a) classification tree analysis (Method 1) and (b) Random forest classification based predictive model (Method 2). Both these methods use student performance in selected assessments during the first two weeks of the semester (Fig. 1).

Method 1(classification tree analysis: manually-created): This method was created based on student performance in formative assessment as a context of student engagement. As noted in [16] [17], incomplete or poor scores in assessment tasks is a sign of disengagement and at risk of dropping out. As such, Method 1 was created to identify students that did not complete or secured low scores in the first two weeks of the semester. students that secured <=25% in selected formative assessments in the year 2018 were classified as at-risk students and visualised for instructors to shepherd, in order to improve student learning outcomes. This student dataset was excluded from predictive analytics (Method 2) (Fig. 1).

Method 2 (Random forest classification based predictive model): This study used Random forest-binary classification model to develop a parsimonious predictive model. Data was collected from 289 students enrolled in an introductory programming course, Introduction to Programming (n=289), over three autumn semesters (years 2016, 2017 and 2018). The initial data collected for model development in the year 2016 was 93. Of these 66 students completed formative assessments and final exam, used as sample data to develop a model with 10-fold cross-validation. The initial data collected for model validation in the year 2017 was 94. Of these 68 students completed formative assessments and final exam, used as validation data to verify the model's performance in line with our research objectives. Similarly, the initial dataset collected for model testing (unknown data) in the year 2018 was 102. The unknown/test dataset, to measure the final model fit for generalisation, was populated from the scores of the 83 students, who scored more than 25% in first two weeks of the semester.

In summary, the data collected via ViLLE [20] for the years 2016 (n=66) and 2017 (n=68) for students who attended the final exam was used to develop and validate a predictive model. The dataset collected for the year 2018 (n=83) was used to test the validated predictive model. That is, the remaining student data (>=25% in formative assessments) for the year 2018 was used to evaluate how well the model developed would perform on unknown data, should instructors wish to use such model for intervention purposes.

### A. *Description of the course and selected formative assessment tasks*

*Introduction to Programming* is offered in English for students with no programming experience. This course covers basic skills in algorithm design, and fundamental concepts of Python programming that includes variables, values, types, expressions, control structures, data structures, modularity and classes. The course duration of 11-12 weeks comprises 24-26 hours of lectures, 16 hours of demonstration sessions, and 10 hours for a practice exam and discussion of project work. This course uses ViLLE, an LMS to deliver lecture materials, formative assessment tasks and to conduct an e-final exam for students [20]. In addition, this tool is used as an automated assessment system to evaluate student-submitted work, to provide immediate feedback for students.

### B. *Formative assessment tasks (FA): Homework (HE) and Demo exercise (DE)*

HE represents weekly continuous formative assessment tasks distributed via ViLLE for students to review material that has been already taught or discussed in class, and to submit their answers electronically over 10-11 weeks. Students are allowed to submit answers for HE multiple times until the due date, with submissions automatically graded via ViLLE. The maximum possible total score for HE is 890.

DE for the course was provided to students weekly for 10 weeks throughout the semester. Programming exercises are provided via ViLLE for students to complete before attending the DE session. A few students are then randomly selected via ViLLE to demonstrate their answers in supervised classes. No marks are awarded for class demonstrations. Students who complete DEs are asked to submit all completed exercises via the lecturer's ViLLE enabled computer. The possible total score for DE is 750.

Both HE and DE are hurdles and must attain at least 50% or more in HE and 40% or more in DE in order to pass these components and to be eligible to sit for final exam. The remaining formative assessment components including project work was included to calculate the final grade for the course. However, project work (assignment) was done in groups of two or three students at the end of the course. Each student in the group had the same final mark allocated for their group work. As such, a mark for project work was not included in the analysis.

### C. *Final exam (FE)*

This is an online summative assessment conducted at the end of the course. This FE is conducted electronically using ViLLE. The FE is a hurdle and student must secure at least 50% to pass the hurdle and the course. The maximum possible score for the FE is 100. The final course grade is calculated based on FE scores, selected formative assessment tasks and lecture attendance scores.

### D. *Final exam Grade (FEG)*

The FEG for the course is calculated based on FE scores. Table I shows the grade calculation used to predict FEG.

TABLE I. GRADE CRITERIA TABLE FOR INTRODUCTION TO PROGRAMMING

| FE marks | Grade* |
|---|---|
| 0 to 49 | 0 (FAIL) |
| 50 to 59 | 1 |
| 60 to 69 | 2 |
| 70 to 79 | 3 |
| 80 to 92 | 4 |
| 93 + | 5 |

*The actual grades 0 and 1 are considered as "at-risk" and denoted as ZERO; Grades 2 to 5 considered as "not-at-risk" and denoted ONE for this study (binary classification)

The formative assessment tasks (HE and DE) data for the course was collected after two weeks of the semesters (2016, 2017, and 2018) for model development, validation and testing. This study used R-software to pre-process the data. The pre-processed data was stored as .xls/csv file to implement the classification tree analysis and predictive modelling. The predictive model was developed with selected formative assessment tasks HE and DE as features to predict student FEG in order to identify at-risk students. This study deployed Random forest classification algorithm with 10-fold cross-validation to predict FEGs. The feature with model equation is HE, DE →FEG.

In this study, we defined students that secured less than 60 marks in FE as at-risk (Table I). This is because; students that secure a passing grade may likely not to succeed in subsequent courses (Tables I & II).

TABLE II. GRADE WISE DISTRIBUTION CALCULATED FROM FE SCORES FOR INTRODUCTION TO PROGRAMMING

| **Final exam grade (FEG)**<br>***At-risk**<br>**** Not-at-risk** | **Number of students** | | |
|---|---|---|---|
| | 2016 | 2017 | 2018 |
| *ZERO = Zero + One | 28 | 16 | 43 |
| **ONE = Two + Three + Four + Five | 38 | 52 | 40 |
| Total | 66 | 68 | 83*** |
| *** Students that secured <=25% in formative assessments in the first two weeks of the semester 2018 were removed from model testing. | | | |

## IV. DATA ANALYSIS AND RESULTS

One of our research methods (Method 1) defined in this study posits that students who underperform in formative assessment tasks in the early weeks of the course will fail in the FE. That is, if the student gets <=25% in HE and DE in the first two weeks of the semester then they may skip or fail in the FE and receive a grade of 0 in the course. Table III & Fig. 2 show the results of the at-risk students who were identified as at-risk based on the decision tree (Method 1) defined in Fig. 1, for the year 2018.

TABLE III. STUDENTS IDENTIFIED AS AT-RISK AFTER TWO WEEKS: METHOD 1 RESULTS

| **<=25 in formative assessments (first two weeks)** | **Did not sit for FE** | **<60 in FE** | **>=60 in FE** | **Accuracy** | **Overall accuracy** |
|---|---|---|---|---|---|
| 19 | 12 | 5 | 2 | 89.5% | 28.3% |

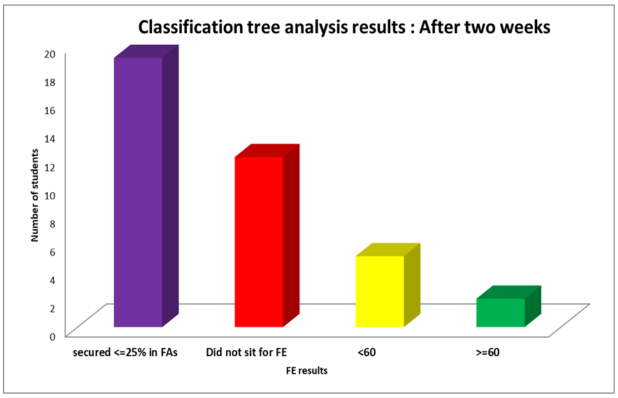

Fig. 2. Method 1 results and performance in FE for the year 2018

In total, 19 students out of 102 that secured <=25% in formative assessments in the first two weeks for the year 2018 were identified as at-risk students and directed for visualisation (Fig. 2). Of those, 12 students did not appear for the FE and received a fail grade. Seven students attended the FE and of those five students secured <60 marks in FE and two students succeeded in the FE despite attaining <=25% in formative assessments in the early two weeks of the semester. In total, 60 students out of 102 secured below 60 marks in the final exam in the year 2018 and received grade 0. Of these, 17 students were correctly identified as at-risk by Method 1. The overall prediction accuracy of Method 1 is 28% [(17/60) * 100] (Table III).

So, we deployed a Random forest classification based predictive model (Method 2) using HE and DE as features to predict FEG in order to identify at-risk students. The classification accuracies of the developed model was evaluated based on confusion matrix computed via R coding. In addition, a performance measurement called receiver operating characteristic curve, referred to as area under the curve was used to evaluate the diagnostic ability of the Random forest classification model developed, validated and tested. The area under the curve is a performance evaluation metric that indicates how well the model is capable of predicting 0s as 0s and 1s as 1s. By analogy, the higher the area under the curve, the better the model is at distinguishing between students with at-risk and not at-risk. The model's overall prediction accuracy was calculated as the number of correct predictions made by the Random forest classification model divided by the total number of actual values, and multiplied by 100 to get the prediction accuracy. Similarly, the at-risk prediction accuracy (sensitivity) was measured by dividing the number of predictions for grade "0" (at-risk) that were correctly identified by dividing the sum of actual number of at-risk students (grade 0) who are correctly identified as they are and the number of at-risk students who are incorrectly identified as not-at-risk students by the model. Then the result is multiplied by 100 to get the at-risk prediction accuracy for the model. Table IV shows the model's overall prediction accuracy and prediction accuracy on identifying students at-risk.

TABLE IV. RANDOM FOREST CLASSIFICATION BASED PREDICTIVE MODEL RESULTS: METHOD 2

| **Dataset & Year** | **Overall prediction accuracy** | **At-risk prediction accuracy** | **Area under the curve** |
|---|---|---|---|
| Training (10-fold): 2016 | 72.73 | 86.84 | 0.73 |
| Validation: 2017 | 52.94 | 37.50 | 0.48 |
| Test set: 2018 | 59.64 | 76.74 | 0.58 |

The Random forest classification model with selected formative assessments (HE & DE) as predictors had significant overall prediction accuracy in training (2016) and moderate in testing (2018) in compliance with area under the curve scores (0.73 & 0.58). In addition, the at-risk prediction accuracy on training and testing in identifying students at-risk of the FE were 54% & 77% respectively. However, the validation data (2017) results on the model had poor overall prediction accuracy and at-risk prediction accuracy (Table IV). This is addressed in discussion section.

As previously noted, 60 students scored a 0 grade (below 60 marks in FE) in the year 2018. Of these, 17 students were identified as at-risk by Method 1 and 33 by Method 2. Therefore by deploying both Methods 1 and 2 it is possible to predict nearly 83% [(17 + 33)/60 * 100] of at-risk students early in the semester.

## V. DISCUSSION

This study proposed the use of two simple prediction methods: classification tree analysis (created manually) and Random forest classification model developed using easily obtainable data sources, that were capable of predicting at-risk students. The ultimate objective of this study was to notify instructors and students who were at-risk of failing the final exam, based on their performances in formative assessments in the first two weeks. First, we analysed the dataset extracted from ViLLE for Introduction to

Programming by using classification tree analysis. Method 1 results (Table III and Fig. 2) suggest that student performance in the early assessment tasks clearly show student engagement and participation in studies. Table III and Fig. 2 clearly reveal that nearly 90% of students (17 out of 19) who secured below <=25% in formative assessments in the year 2018 received grade 0 in FE. Table III suggest that students who did poorly in formative assessments in the early weeks may not sit or fail in the FE. These results also imply that it is possible to identify at-risk learners after week 2, based on student performance in selected formative assessments. In addition, these results support the learning edge momentum effect in computer science discussed by Robbins [47], that not acquiring the concepts in the first few weeks may result in difficulties in learning other linked concepts acquiring in subsequent weeks, as concepts in programming are tightly integrated.

Towards the aforementioned objective, a simple predictive model was developed, with a combination of minimum number of explanatory variables selected based on prior study [46] which are easily obtained early in the term or semester by instructors to predict students at risk. We used two continuous formative assessment tasks; (i) DE: partially assisted by instructor in the classroom and (ii) HE: take-home assessment task as predictors to build a model in order to predict student FEG. The overall prediction accuracy and at-risk prediction accuracy of Random forest classification based predictive model developed, validated, and tested with selected formative assessments produced mixed results. The overall prediction accuracy and at-risk prediction accuracy results on training set (year 2016: 73% & 54%) and unknown dataset (year 2018: 60% & 77%) were significant in compliance with area under the curve scores (0.73 & 0.58) and suggest that student performance in formative assessment in the early weeks may be appropriate to identify students at-risk very early (Table IV). On the other hand, the overall prediction accuracy and at-risk prediction accuracy results for the validation set (year 2017) were insignificant and prompted a drilling down of the data used for validation. The post analysis on the 2017 dataset revealed that there was no significant variation among student demographics, course periods, assignments, exams, and instructor. However, it seems, the structure of the 2017 dataset is skewed (Table II). That is, the distribution of at-risk and not-at-risk classes of the 2017 data are not equally distributed in comparison to other datasets used for training (year 2016) and testing (year 2018), although unknown dataset results of our model support the use of formative assessments as early warning indicators. Hence, it should be noted that the model developed with insufficient or imbalanced data may affect the predictive performance of the model [48]. We also analysed student prior programming knowledge using data collected via a course entry survey conducted at the beginning of the course. Sixty percent of students who enrolled in 2017 had no prior knowledge in programming. On average, only 44% students who enrolled in 2016 and 2018 had no prior knowledge in programming. Hence, we surmised that, the structure of the dataset and student prior knowledge in programming may have influenced our predictive model results. This needs further analysis. These prediction results are in congruent with the results obtained in the study of Liao et al. [37]. That is, it should not be assumed that courses will not change and, moreover, prediction methods should be flexible enough to adapt to the realities of teaching.

In addition, from these results the following points emerged. First, formative assessments are designed to educate and improve student problem-solving skills and performance. However, if a student is not genuinely involved or has not developed a deep learning via formative assessments, then it is difficult to perform well/better in the FE. Second, it is also possible that students might have done formative assessments by using resources such as the internet and other support systems for solution ideas. However, these need further analysis. Despite these mixed results, our classification tree analysis and Random forest classification based predictive model results support the importance of formative assessment in identifying student learning and FE performance. In addition, these results imply that formative assessment results in the first two weeks may serve as early warning indicators for identifying students at-risk of failing or not attending the FE.

Similarly, the results for the year 2016 and 2018 raise the following questions: “How a research methodology developed for this study can be deployed for other programming and non-programming courses if the goal of instructor is to identify students who will skip, fail or marginally succeed in the course?” Or “How these results could be used to further support identified at-risk students?” As we concluded earlier, early assessments can be considered as a possible predictor for predictive models to gauge student progress early in the course that has formative and summative assessments. The methods and results from this analysis can be used by instructors and students to provide early interventions before they disengaged from studies. For example, the results drawn from the decision tree analysis & Random forest classification predictive model defined for this study may be used by instructors to categorize students as “at-risk”, “marginal”, “good” based on the formative assessment scores achieved in every two weeks of the course in order to reshape pedagogical practices accordingly. After identifying low performers of formative assessment via classification tree and Random forest classification predictive model, instructors may counsel those students about strategies to improve their performance. Similarly, students may also use our developed methods in the following ways: First, the student continuous formative assessment scores may be delivered as real-time feedback to identified at-risk students to push them to devote attention on learning activities to improve learning performance. Second, students may perceive these results as ongoing performance indicators to alter their academic behaviours, and to understand their performance, shortcomings and their success.

Finally, critical to this study is the question: “How visualising patterns of student ongoing assessment performance assist stakeholders to improve student outcomes?” Visualising student educational activity data captured by a learning management system or e-learning tool for stakeholders is an important reporting function in learning analytics. Reporting is the second stage of a learning analytics engine [12]. Therefore, visualising student performance data would potentially assist instructors to learn and quickly identify disengaged learners. In addition, visualising formative assessment performance with classification of performance level would help instructors to extract patterns of performance, areas of weakness or strength, and to identify students who need more attention than others. Similarly, presenting ongoing assessment performance to students in visual form, with recommendations and feedback, would potentially assist students to increase self-awareness of their participation in a course. For example, in this study we tracked and displayed the students based on formative assessment performance in the first two weeks for instructors to detect students at risk of abandoning or failing the course (Table III and Fig. 2). This process can be replicated. That is, student performance in formative assessments and the results drawn from Method 1 and Method 2 may be visualised at least once every two weeks of the course as a performance activity monitor for instructors and students. Moreover, monitoring performance of students who need support would assist academic advisors and instructors to request/extract additional data from the course management systems, in order to provide improved teaching interventions in a timely manner. However, the impact of this kind of reporting should be examined to know how students will react to such reporting. Feedback provided by instructors to students should not discourage them from learning. Thus, the challenge for instructors or stakeholders is to have a clear policy framework for designing and deploying visualisation of student engagement and performance data. Students need to receive timely and specific feedback to return to their learning strategies, to increase learning outcomes.

## VI. CONCLUSIONS, LIMITATIONS AND FUTURE WORK

Our classification tree analysis and predictive modelling based study showed that results of early assessments might be used as early warning signs to identify students who need support. The Method 1 results showed that, early formative assessment scores (homework and demo exercises) presents a sign of student disengagement in studies. Moreover, statistical results suggest that students who underperformed in formative assessment in the early weeks of the course might not attend the FE or receive a failing grade (Table III). The results of this study provide evidence that formative assessment partially impacts student achievement. Similarly, the Random forest classification based model results (Table IV) suggest that, it is possible to predict and identify at-risk students during their first few weeks of the course.

This research has few a limitations. First, the results of this study were derived from both traditional classification tree analysis and Random forest classification based predictive model and applied to a specific university course. Second, the sample size of the study was not sufficiently large to generalize our findings. Third, we used the first two weeks of assessment results for analysis. However, learning is dynamic and a learner might not do well in the beginning weeks of the semester and may perform well in subsequent weeks of the semester. Hence, there is a need to monitor and track student progress throughout the course period in order to provide continuous academic support. Fourth, this study did not compare the results of the methods presented, to gauge differences between two independent groups. This will be considered in a future study. Fifth, this study did not focus on other causal factors such as student prior knowledge in programming, problem solving skills, prior GPA, motivation, and self-esteem. Finally, although the accuracy in identifying at-risk students on test data was good (Tables III and IV); it is unknown weather this approach might work on unknown data, as the validation results of the model were not significant.

Despite these limitations, our findings provide some suggestions for learners and instructors, course administrators and academic advisors. Formative assessment performance strongly related to student performance and implies that it has substantial impact on the learning process. Therefore, we contend that our developed classification tree and predictive model techniques might be applied to other courses with continuous formative assessments and a FE to identify students that need support early. Moreover, our analysis may be extended to seek answers to the following research questions: (i) How does student formative assessment performance influence student learning? (ii) What kind of support activities might be provided to at-risk students identified by this classification analysis? Indeed, of interest in future work, the developed classification tree and predictive model might be extended with more decisions/criteria incorporating additional factors such as number of assignments submitted, the number of submissions made, and the differences between first submission score and subsequent submission scores.

## ACKNOWLEDGMENT

The authors wish to thank all ViLLE team research members, Department of Future Technologies, University of Turku for their comments and support that greatly improved the manuscript.

## REFERENCES

[1] C. Watson and F. W. Li, "Failure Rates in Introductory Programming Revisited," in Proceedings of the 2014 conference on Innovation & technology in computer science education, Uppsala, 2014.

[2] J. Bennedsen and M. E. Caspersen, "Failure Rates in Introductory Programming — 12 Years Later," ACM inroads, vol. 10, no. 2, June 2019.

[3] G. S. Abu-Oda and A. M. El-Halees, “Data Mining in Higher Education: University student dropout case study,” International Journal of Data Mining & Knowledge Management process, vol. 5, no. 1, pp. 15-27, 2015.

[4] K. Longi, "Exploring factors that affect performance on introductory programming courses," Helsinki, 2016.

[5] T. Agasisti and A. J. Bowers, "Data Analytics and Decision Making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions," in Handbook of Contemporary Education Economics, G. J. J. A. T. L. L. Johnes, Ed., Cheltenham, Edward Elgar Publishing, 2017, pp. 184-210.

[6] A. J. Bowers and X. Zhou, "Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes," Journal of Education for Students Placed At Risk, vol. 24, no. 1, pp. 20-46, 2019.

[7] A. Jokhan, B. Sharma and S. Singh, "Early warning system as a predictor for student performance in higher education blended courses," Studies in Higher Education, pp. 1-12, 22 May 2018.

[8] A. E. Krumm, R. J. Waddington, S. D. Teasley and S. Lonn, “A learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering,” in Learning Analytics: From Research to Practice, J. A. Larusson and B. White, Eds., New York, Springer, 2014, pp. 103-119.

[9] R. Asif, A. Merceron and M. K. Pathan, “Predicting Student Academic Performance at Degree Level: A Case Study,” International Journal of Intelligent Systems and Applications, vol. 7, no. 1, pp. 49-61, 2015.

[10] Z. G. Baleni, "Online formative assessment in higher education: Its pros and cons," The Electronic journal of e-Learning, vol. 13, no. 4, pp. 228-236, 2015.

[11] G. J. Cizek, H. L. Andrade and R. E. Bennett, "Formative Assessment : History, Definition, and Progress," in Handbook of Formative Assessment in the Disciplines, R. E. B. G. J. C. Heidi L. Andrade, Ed., New York, Routledge, 2019, pp. 1-17.

[12] A. Pardo, "Designing Learning Analytics Experiences," in Learning Analytics From Research to Practice, J. Larusson and B. White, Eds., New York, Springer, 2014, pp. 15-35.

[13] P. Jia and T. Maloney, "Using predictive modelling to identify students at risk of poor university outcomes," Higher Education, vol. 70, no. 1, pp. 127-149, 2015.

[14] J. Horton, "Identifying At-Risk Factors That Affect College Student Success," International Journal of Process Education, vol. 7, no. 1, pp. 83-102, June 2016.

[15] High School Dropout, Graduation, and Completion, Washington DC: National Academies Press, 2011.

[16] B. A. Boyd, "Early-Warning Indicators of High School Dropout," OhioLINK Electronic Theses and Dissertations Center, Columbus, 2016.

[17] M. Bruce, J. M. Bridgeland, J. H. Fox and R. Balfanz, "On Track for Success: The Use of Early Warning Indicator and Intervention Systems to Build a Grad Nation," Civic Enterprises, Washington, 2011.

[18] L. Chipchase, M. Davidson, F. Blackstock, R. Bye, P. Clothier, N. Klupp, W. Nickson, D. Turner and M. Williams, "Conceptualising and Measuring Student Disengagement in Higher Education: A Synthesis of the Literature," International Journal of Higher Education, vol. 6, no. 2, pp. 31-42, 2017.

[19] D. M. Horton, S. D. Wiederman and D. A. Saint, "Assessment outcome is weakly correlated with lecture attendance: influence of learning style and use of alternative materials," Advances in Physiology Education, vol. 36, no. 2, pp. 108-115, 2012.

[20] Veerasamy, A. K., D'Souza, D., Lindén, R., Kaila, E., Laakso, M.-J., & Salakoski, T. (2016). The Impact of Lecture Attendance on Exams for Novice Programming Students. International Journal of Modern Education and Computer Science (IJMECS), 8(5), 1-11. doi:10.5815/ijmecs.2016.05.01.

[21] E. Trumbull and A. Lash, Understanding Formative Assessment: Insights from Learning Theory and Measurement Theory, San Francisco: WestEd, 2013.

[22] I. Gratchev and A. Balasubramaniam, "Developing assessment tasks to improve the performance of engineering students," in 23rd Annual Conference of the Australasian Association for Engineering Education 2012, Melbourne, 2012.

[23] M. T. Carrillo-de-la-Pena, E. Bailles, X. Caseras, A. Martinez, G. Ortet and J. Perez, "Formative assessment and academic achievement in pre-graduate students of health sciences," Adv in Health Science Education, vol. 14, no. 1, pp. 61-67, 2009.

[24] H. Fan, J. Xu, Z. Cai, J. He and X. Fan, "Homework and students' achievement in math and science: A 30-year meta-analysis, 1986–2015," Educational Research Review, vol. 20, pp. 35-54, 2017.

[25] G. Baş and C. Ş. a. F. M. Ciğerci, "Homework and academic achievement: A meta-analytic review of research," Issues in Educational Research, vol. 27, no. 1, pp. 31-50, 2017.

[26] F. V. Gaal and A. D. Ridder, "The impact of assessment tasks on subsequent examination performance," Active Learning in Higher Education, vol. 14, no. 3, pp. 213-225, 2013.

[27] A. Y. S. Su, C. S. J. Huang, S. J. H. Yang, T. J. Ding and Y. Z. Hsieh, "Effects of Annotations and Homework on Learning Achievement: An Empirical Study of Scratch Programming Pedagogy," Journal of Educational Technology & Society, vol. 18, no. 4, pp. 331-343, 2015 October 2015.

[28] W. Li, R. M. Bennett, T. Olsen and R. McCord, "Engage Engineering Students in Homework: Attribution of Low Completion and Suggestions for Interventions," American Journal of Engineering Education, vol. 9, no. 1, pp. 23-38, 2018.

[29] T. Devasia, V. T. P and V. Hegde, "Prediction of students performance using Educational Data Mining," in 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), Ernakulam, 2016.

[30] K. Castro-Wunsch, A. Ahadi and A. Petersen, "Evaluating Neural Networks as a Method for Identifying Students in Need of Assistance," in In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, Seattle, 2017.

[31] R. Fujinuma and L. A. Wendling, "Repeating Knowledge Application Practice to Improve Student Performance in a Large, Introductory Science Course," International Journal of Science Education, vol. 37, no. 17, pp. 2906-2922, November 2015.

[32] P. B. W. H. a. A. T. Jens Dolin, "Exploring Relations Between Formative and Summative Assessment," in Transforming Assessment : Through an Interplay Between Practice, Research and Policy, Springer, 2018, pp. 53-80.

[33] A. Ahadi, R. Lister, H. Haapala and A. Vihavainen, "Exploring Machine Learning Methods to Automatically Identify Students in Need of Assistance," in Proceedings of the eleventh annual International Conference on International Computing Education Research, Omaha, Nebraska, USA, 2015.

[34] L. Porter, D. Zingaro and R. Lister, "Predicting student success using fine grain clicker data," in Proceedings of the tenth annual conference on International computing education research, Glasgow, Scotland, United Kingdom, 2014.

[35] S. N. Liao, D. Zingaro, M. A. Laurenzano, W. G. Griswold and L. Porter, "Lightweight, Early Identification of At-Risk CS1 Students," in Proceedings of the 2016 ACM Conference on International Computing Education Research, Melbourne, VIC, Australia, 2016.

[36] K. Quille and S. Bergin, "Programming: predicting student success early in CS1. a re-validation and replication study," in Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, Larnaca, Cyprus, 2018.

[37] S. N. Liao, D. Zingaro, K. Thai, C. Alvarado, W. G. Griswold and L. Porter, "A Robust Machine Learning Technique to Predict Low-performing Students," ACM Transactions on Computing Education (TOCE), vol. 19, no. 3, pp. 18:1 -18:19, June 2019.

[38] S. N. Liao, D. Zingaro, C. Alvarado, W. G. Griswold and L. Porter, "Exploring the Value of Different Data Sources for Predicting Student Performance in Multiple CS Courses," in Proceedings of the 50th ACM Technical Symposium on Computer Science Education, Minneapolis, MN, USA, 2019.

[39] P. Oskar Gislason, J. A. Benediktsson and J. R. Sveinsson, "Random Forests for land cover classification," Pattern Recognition Letters, vol. 27, no. 4, 2006.

[40] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in Proceedings of 5th Annual Future Business Technology Conference, Porto, 2008.

[41] K. P. Rao, M. C. Sekhara and B. Ramesh, "Predicting Learning Behavior of Students using Classification Techniques," International Journal of Computer Applications, vol. 139, no. 7, pp. 15-19, April 2016.

[42] M. Kumar and A. Singh, "Evaluation of Data Mining Techniques for Predicting Student's Performance," International Journal of Modern Education and Computer Science, vol. 8, pp. 25-31, 2017.

[43] S. Hussain, N. A. Dahan, F. M. Ba-Alwib and N. Ribata, "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA," Indonesian Journal of Electrical Engineering and Computer Science, vol. 9, no. 2, pp. 447-459, 2018.

[44] B. Perez, C. Castellanos and D. Correa, "Predicting Student Drop-Out Rates Using Data Mining Techniques: A Case Study," in IEEE Colombian Conference on Applications in Computational Intelligence, 2018.

[45] H. Bydžovská, "A Comparative Analysis of Techniques for Predicting Student Performance," in Proceedings of the 9th International Conference on Educational Data Mining, Raleigh, NC, USA.

[46] Veerasamy, A. K., D'Souza, D., Lindén, R., & Laakso, M.-J. (2019, February 1). Prediction of Student Final Exam Performance in an Introductory Programming Course: Development and Validation of the Use of a Support Vector Machine-Regression Model. Asian Journal of Education and E-learning, 7(1), 1-14.

[47] A. Robins, "Learning edge momentum: a new account of outcomes in CS1," Computer Science Education, vol. 20, no. 1, pp. 37-71, 7 April 2010.

[48] F. Marbouti, H. A. Diefes-Dux and K. Madhavan, “Models for early prediction of at-risk students in a course using standards-based grading,” Computers & Education, vol. 103, pp. 1-15, 2016.