Research, Society and Development, v. 11, n. 5, e6511527938, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i5.27938

The application of Models Portability to predict undergraduate students' performance by using Transfer Learning

Aplicação de Portabilidade de Modelos para predição de desempenho de estudantes de graduação usando Transfer Learning

Aplicación de Portabilidad de Modelos para predicción de desempeño de estudiantes de pregrado usando Transferencia de Aprendizaje

Received: 03/10/2022 | Reviewed: 03/18/2022 | Accept: 03/23/2022 | Published: 03/29/2022

Carlos Antonio R. Beltran
ORCID: https://orcid.org/0000-0002-7947-1352
Federal University of Rio Grande do Norte, Brazil
E-mail: carvan1521@gmail.com
João Carlos Xavier Júnior
ORCID: https://orcid.org/0000-0003-1517-2211
Federal University of Rio Grande do Norte, Brazil
E-mail: jcxavier@imd.ufrn.br
Cephas Alves da Silveira Barreto
ORCID: https://orcid.org/0000-0002-4756-8571
Federal University of Rio Grande do Norte, Brazil
E-mail: cephasax@gmail.com
Arthur Costa Gorgônio
ORCID: https://orcid.org/0000-0002-1824-9600
Federal University of Rio Grande do Norte, Brazil
E-mail: gorgonioarthur@gmail.com
Song Jong Márcio Simioni da Costa
ORCID: https://orcid.org/0000-0002-1980-7969
Federal University of Rio Grande do Norte, Brazil
E-mail: marcio.simioni.017@ufrn.edu.br

Abstract

One of the great challenges of education in recent years has been to accurately and reliably predict students' performance in order to apply different strategies in order to help them with their academic deficiencies. Based on this fact, the main goal of this work is to apply a Transfer Learning approach on Learning Management Systems logs (i.e., Moodle) in order to achieve good portability of models and then predict the performance of undergraduate students. Two different scenarios have been implemented considering the activities of each course used in Moodle, the group of similar courses of the same degree as the first scenario and the group of a similar level of usage of activities as the second one. Empirical analysis has been conducted in order to evaluate the performance of the models created with three well-known classification algorithms (i.e., Decision Tree, Random Forest and Naive Bayes). AUC ROC, F-Measure, Precision and Recall have been applied as prediction measures for choosing the best models and evaluating their portability performance to the other courses. Even in the early stage, the experimental results encourage us to state that it is possible to apply transfer predictive models to the same group of courses in the majority of the cases.

Keywords: Transfer learning; Machine learning; Student performance; Moodle.

Resumo

Um dos grandes desafios da educação nos últimos anos tem sido prever com precisão e confiabilidade o desempenho dos alunos a fim de aplicar diferentes estratégias para ajudá-los em suas deficiências acadêmicas. Com base neste fato, o objetivo principal deste trabalho é aplicar uma abordagem de Transferência de Aprendizagem em logs de Sistemas de Gestão de Aprendizagem (i.e., Moodle) a fim de obter uma boa portabilidade de modelos e, com isso, prever o desempenho dos alunos de graduação. Dois cenários diferentes foram implementados considerando as atividades de cada curso utilizado no Moodle, o primeiro cenário, com o grupo de cursos similares de mesma graduação, e o segundo cenário, com o grupo de níveis de utilização de atividades. A análise empírica foi realizada para avaliar o desempenho dos modelos criados com três algoritmos de classificação bem conhecidos (i.e., Árvore de Decisão, Random Forest e Naive Bayes). Além disso, as métricas AUC ROC, F-Measure, Precision e Recall foram usadas como medidas de predição para escolher os melhores modelos e avaliar seu desempenho de portabilidade para os demais cursos. Os resultados experimentais nos encorajam a afirmar que é possível aplicar a transferência de modelos preditivos para o mesmo grupo de cursos na maioria dos casos.

Palavras-chave: Transferência de aprendizado; Aprendizado de Máquina; Desempenho do aluno; Moodle.

Resumen

Uno de los grandes retos de la educación en los últimos años ha sido predecir con precisión y fiabilidad el rendimiento de los alumnos para poder aplicar distintas estrategias que les ayuden a afrontar sus deficiencias académicas. Basado en este hecho, el objetivo principal de este trabajo es aplicar un enfoque de transferencia de aprendizaje en los registros del sistema de gestión de aprendizaje (i.e., Moodle) para obtener una buena portabilidad del modelo y, con eso, predecir el rendimiento de los estudiantes de pregrado. Se implementaron dos escenarios diferentes considerando las actividades de cada curso utilizado en Moodle, el primer escenario, con el grupo de cursos similares de la misma especialidad, y el segundo escenario, con el grupo de niveles de uso de actividades. Se realizó un análisis empírico para evaluar el rendimiento de los modelos creados con tres algoritmos de clasificación bien conocidos (i.e., Árbol de Decisión, Bosque Aleatorio y Naive Bayes). Además, las métricas AUC ROC, F-Measure, Precision y Recall se utilizaron como medidas predictivas para elegir los mejores modelos y evaluar su rendimiento de portabilidad a los otros cursos. Los resultados experimentales nos animan a afirmar que es posible aplicar la transferencia de modelos predictivos a un mismo grupo de cursos en la mayoría de los casos.

Palabras clave: Transferencia de aprendizaje; Aprendizaje automático; Rendimiento estudiantil; Moodle.

1. Introduction

Higher education is an indispensable requirement for the growth and development of a country since, by training several professionals and promoting quality research, it can become a backbone for the nation to develop economically strong and globally influential. However, many Higher Education Institutions often have to deal with problems such as economic uncertainty, lack of quality research, lack of cooperation with the industrial sector and poor academic performance (Palma et al., 2011).

In this sense, focusing on these problems, it is important to emphasise that the main reasons for the poor academic performance of higher education students are usually related to multiple factors, including, for example, family problems, parents' low income, social and cultural environment. The more these factors remain unresolved, the more they will affect the students' bio and psychological integrity. As a consequence, students may develop problems, including attention deficit, poor learning mentality, low enthusiasm for learning and lack of motivation for learning, which will result in low productivity and, therefore, poor academic performance (Ayala & Manzano, 2018).

It is also important to clarify the concept of academic performance. In educational terms, performance is one of the learning outcomes generated by the teacher's educational activities and that has an effect on all students, even though not all learning is a product of the teacher's action (Davidson, 2002). Academic performance is related to factors such as intelligence and ability. According to Fernandez-Berrocal e Checa (2016), intelligence refers to the cognitive abilities that enable the individual to perform different activities and to adapt them to the demands of the environment. In this sense, some learning tools have been used to support the learning process by providing online activities for higher education courses (Ulker & Yilmaz, 2016).

By adopting these learning tools, in particular Learning Management Systems (LMS), higher educational institutions are able to analyse enormous quantities of data that describe the behaviour of students. Among these important tools, which can add web technology to courses and supplement the traditional face-to-face approaches, we can point out Moodle (Dougiamas & Taylor, 2003) as being the most popular LMS. Moodle accumulates a great amount of information which is very valuable for mining students' behaviour in the educational domain. It keeps detailed logs of all events that students perform and keeps track of what materials students have accessed.

In general, Moodle logs are used by Educational Data Mining (EDM) and Learning Analytics (LA) tools (Romero et al., 2008). Although EDM and LA techniques are able to discover useful knowledge from educational data, they usually aim to predict students' performance by estimating the unknown value of variables that can describe their performance (e.g., knowledge, score, or mark) based on scenarios that use training and test data from the same course (Romero & Ventura, 2013). Arguably, the ideal scenario will be the one that could have models obtained over data from particular courses but able to be used for different ones.

In this sense, some research has been conducted in order to use the idea of model portability, which is based on creating and using transferable models obtained over similar course data. The idea of portability is that knowledge extracted from a specific course can be applied to another different course (López-Zambrano et al., 2020). Most of the works related to model portability use a Transfer Learning (TL) approach in which an obtained model is transferred from one course to another (Boyer & Veeramachaneni, 2015). Moreover, these predictive models reported in the existing studies try to predict whether a student will pass or fail by using the log data generated from the student interactions with Moodle LMS. However, few works have deeply analysed different Machine Learning techniques and, more importantly, the variety of courses available in universities.

In this context, the main contribution of this work is to analyse the portability of models by using different interpretable Machine Learning techniques over Moodle data logs from different university courses. In general, most of the related work uses only one type of technique and one performance metric. Our goal is to perform a robust comparison analysis on the portability of trained models to be used over different course data from the Catholic University Los Angeles of Chimbote in Peru.

The remainder of this paper is organised as follows. Section 2 discusses background on Learning Management System (LMS) and Supervised Learning (SL). Section 3 discusses related work on Transfer Learning. Section 4 describes the Experimental Methodology. Section 5 presents the computational results. Finally, Section 6 presents our conclusions and a direction for future work.

2. Background

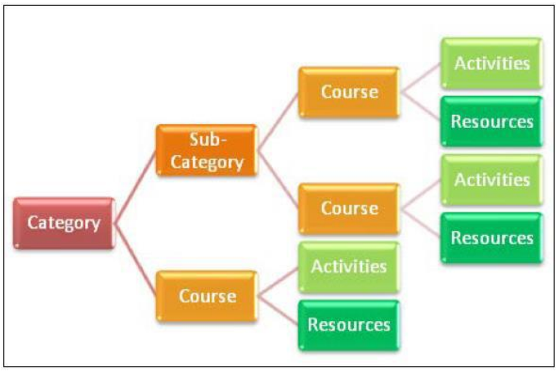
2.1 Learning Management System (LMS)

LMSs are tools that provide an educational institution with the ability to train, manage, teach and monitor students. Due to the rapid recognition of the academic benefits of online training costs, LMSs have grown considerably in recent years (Piña, 2012). There are a few LMS tools such as Blackboard, Desire2Learn, Angel, Chamilo, Claroline and Moodle. All these tools have become support for higher education, as they are widely used in the teaching structure of colleges and universities. In this work, the LMS Moodle will be used as a platform with a free Open Source licence, which allows anyone to contribute to its development or improvement.

Moodle offers a constructive, interactive, integrated and learning-centred environment. The basic principle that guides this platform is social constructivism, which implies collaborative learning based on individual and group projects and tasks. Thus, learning becomes effective when the teacher builds a learning material capable of making their students interact and learn through it (Raga & Raga, 2017).

The interaction between teacher and students with Moodle generates a large amount of data logs. Logs in Moodle can be divided in categories, courses, activities, and resources. Within a course, different types of activities can be offered, such as: Assignment, Chat, Survey, Forum, Quiz, Lesson, Assessment Survey and Wiki. Moreover, Moodle stores learning resources (e.g., folder, page and link) to help activities in courses. On top of that, categories act as containers for courses. They can have subcategories, which can have sub-subcategories, and so on. This hierarchical structure can be visualised as follows (Büchner, 2016):

Figure 1. Moodle's Hierarchical structure.



Source: Büchner (2016).

2.2 Supervised Learning (SL)

In the classification task of Machine Learning (ML), each instance in the input dataset is represented by a set of features and a class attribute. With access to the class values of instances in the training set, but not in the test set, a classification algorithm has the goal to learn a model from the training set that is able to predict the class value of each instance in the test set (with instances unseen during training), based on the feature values for that instance (Barros et al., 2015).

Many types of classification algorithms have been proposed and are often used, such as Decision Tree, Neural Networks, Support Vector Machine, Random Forest, Naive Bayes, among many others. However, in general, different types of classification algorithms have different pros and cons, and different biases; therefore, no single type of classification algorithm can be considered the best for all datasets or application domains Fernández-Delgado et al., (2014). In this sense, as model interpretation is vital, we decided to use classification algorithms with reasonable interpretation such as: Decision Tree, Naive Bayes (Han & Kamber, 2011) and Random Forest (Breiman, 2001).

2.2.1 Predictive Accuracy Measures

The development of predictive models and algorithms demands the need of quantifying how robust the models are to future observations. Quantifying the accuracy of an algorithm is an important issue, and for this reason needs to consider the characteristics of the data. In general, the available datasets are imbalanced, thus it is necessary to use predictive accuracy measures that are able to take in consideration this nature of the data itself. Among the predictive accuracy measures found in Literature, we will use Precision, Recall, F-Measure and Area Under the Curve (AUC) ROC.

Precision is the ratio of the number of correctly classified positive instances over the total number of instances classified as positive (regardless of whether they belong to the positive or negative class). Recall is the ratio of the number of correctly classified positive instances over the total number of positive instances (regardless of them being correctly or wrongly classified). The F-measure is the harmonic average between precision (prec) and recall (rec), defined as:

$$F\text{-}Measure = \frac{2 \times precision \times recall}{precision + recall}$$

where, Precision and Recall can be defined as follows:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall \ = \frac{TruePositive}{TruePositive \ + \ FalseNegative}$$

The Receiver Operating Characteristic (ROC) curve is a graph plotting the pair of (1 − specificity, sensitivity) for all possible threshold values, demonstrating a trade-off, phenomena, between sensitivity and specificity. The area under the ROC curve (AUC) is a very popular formal index used to summarise the ROC curve. Therefore, AUC ROC can be defined as follows:

$$AUC\ ROC = \int_{0}^{1} \ \ TPRate\ \partial(FPRate)$$

where, TP Rate and FP Rate can be defined as follows:

$$TPRate \ = \frac{TruePositive}{TruePositive \ + \ FalseNegative}$$

$$FPRate \ = \frac{FalsePositive}{FalsePositive \ + \ TrueNegative}$$

2.3 Transfer Learning (TL)

Traditional supervised machine learning tasks involve training and testing datasets having the same input feature space and data distribution. However, as it has already been proved by some works (Weiss et al., 2016; Hunt et al., 2017; Pan & Yang, 2009; Sarkar et al., 2018), in case of existing difference in data distribution between the training (source domain) and test (target domain) sets, the results of a predictive learner can be degraded. Based on this fact, there is an urgent need to create high-performance learners trained from related source domains but applicable to different target domains.

Transfer Learning is used to improve the efficiency of a learner by transferring information from one domain to a related one. As TL has been formally defined in Weiss et al., (2016) and Pan e Yang (2009), we can assume that given a source domain D_S with a corresponding source task T_S, and a target domain D_T with a corresponding task T_T, transfer learning is the process of improving the target predictive function $f_T(.)$ by using the related information from D_S and T_S, where $D_S \neq D_T$ or $T_S \neq T_T$.

3. Related Work

In the last years Machine Learning has been used as a way to solve real problems in a wide range of fields, such as medicine (Teles et al., 2021), image recognition (Neves et al., 2019), pharmaceutics (Mangini et al., 2021), marketing and gastronomy (Ossani et al., 2021), education (Mascarenhas et al., 2020), and others. In this sense, educational problems, such as predicting the students' performance have received special attention. One of the research efforts in the academic field is related to the use of transfer learning to promote the reusing of ML models built based on datasets from different contexts and courses. This section reviews related works on the prediction of academic performance of higher education students by using Machine Learning (3.1) and Transfer Learning (3.2) approaches. Note that both subsections focus on higher education tasks related to either students’ performance or dropout.

3.1 Prediction of Students' Performance by Using Machine Learning Approaches

Current researches on prediction of academic performance of higher education students apply supervised and unsupervised ML techniques over datalogs from LMSs (Raga & Raga, 2017; Nguyen et al., 2018; Aleksandrova, 2019; Quinn & Gray, 2020; Ahmed et al., 2021; and Hao et al., 2022). In Raga e Raga (2017), for instance, the authors proposed a Vector Space Model approach for processing and summarising students' activities from Moodle's datalogs. This approach resulted in a vector-based form used to map students' activity in a latent activity space given a set of activity dimensions (i.e., resulted in a flat dataset). Thus, they used classification algorithms (e.g., Logistic Regression, Classification and Regression Trees, Random Forest, BayesNet and k-NN) over this dataset. Experiments indicated that the generated model could distinguish between sets of activities that lead to High, Low, or Failed performances.

In the same context, Quinn e Gray (2020) investigated the applicability of data logs from Moodle to be used for predicting students' academic performance in a blended learning further education setting. The goal was to predict alphabetic student grades, and whether a student would pass or fail the course. According to them, classifiers built on all course data predicted student grade moderately well (accuracy = 60.5%) and whether a student would pass or fail very well (accuracy= 92.2%). However, classifiers built on the first six weeks of data did not predict failing students well. Classification algorithms used in these experiments were Random Forest (RF), Gradient Boosting, K-Nearest Neighbors (k-NN) e Linear Discriminant Analysis (LDA).

Moreover, Aleksandrova (2019) focused on the application of supervised machine learning algorithms for predicting students' performance based on their interaction with Moodle's data logs. In this analysis, several classifiers were used, such as: Logistic Regression, Random Forest, Gradient boosting decision trees (XG boost) and Neural network. The results showed that the classifiers performed satisfactorily with accuracy above 84% and that neural network and XG boost outperformed the other two classifiers.

On the other hand, Nguyen et al. (2018) proposed a forecast learning outcome model based on learner interaction with online learning systems by providing a learning analytics dashboard to learners and teachers for monitoring and online orientation of learners based on some machine learning and data mining techniques. Their goal was to analyse the feasibility of predicting learners' learning outcomes based on their interactive activities and the best way to monitor and guide learners in an effective online learning environment. Differently from the three previous works, the authors applied Clustering algorithms (e.g., K-means, Birch, and Hierarchical Agglomerative) in order to select the more suitable partitions to be used in learning analytics for predicting learning outcomes of learners through learning activities.

In Ahmed et al., (2021), the authors used the Gradient Boosting Decision Tree (GBDT) method to predict the students' performance in final exams. The authors compared the proposal against several ML methods, such as Support Vector Machine, Logistic Regression and Naive Bayes, and the results were promising. A more elaborated system was proposed in Hao et al., (2022), in which the authors used a Bayesian Network to compose a system named Students' performance Prediction Bayesian Network (SPBN). This system used hill-climbing and maximum likelihood estimation (MLE) to predict the students' performance and a Genetic Algorithm (GA) to give personalised improvement suggestions for students about to fail the courses. The experiments used the Open University Learning Analytics Dataset (OULAD), and the results showed that the proposed approach obtained high prediction performance and gave reasonable improvement suggestions.

Finally, all related work reported here (subsection 3.1) focused on building the best classifier to predict students' performance by using the entire course dataset, or applying clustering analysis in order to select the best partitions to be used in learning analytics. In the case of the first approach, training is always requested for a new dataset, which invalidates the idea of having a portable trained model. Whereas in the second case, clustering techniques might produce results completely different from one dataset to another, even with the same features.

3.2 Prediction of Students' Performance by Using Transfer Learning Approaches

In scenarios where high-performance learners trained with data from different domains are needed, the traditional Machine learning techniques are no longer efficient. Therefore, Transfer Learning (TL) has been applied to real-world applications with these needs. Weiss et al. (2016), presents a complete survey on TL, discussing information on current solutions and reviews applications applied to it. However, as this work focus on TL applied to predict the academic performance of higher education students, we will report only works related to this specific area, and that can be found in literature, such as in Boyer e Veeramachaneni (2015); Hunt et al., (2017); Ding et al., (2019); López-Zambrano et al., (2020); Tsiakmaki et al., (2020); and López-Zambrano et al., (2021).

Data recorded from Massive Open Online Courses (MOOC) platforms train learners to provide an excellent opportunity to build predictive models that can help anticipate future behaviours and develop interventions. In Boyer & Veeramachaneni (2015), for instance, the authors used data from previous courses and the first weeks of the same course to make real-time predictions on learners' behaviour (i.e., the stop out prediction problem). In this work, they evaluated multiple transfer learning methods in order to estimate the feasibility of transferring knowledge across courses.

Similarly to previous work, Hunt et al., (2017) proposed an approach for predicting graduation rates in degree programs by using data from diverse programs. Initially, information from multiple degree programs was retrieved to construct an effective sample size. Then a process took into account the differences across several degree programs and automatically down-weighted less-relevant data. Finally, a model was built over this data, allowing the knowledge to be transferred from a more vast domain to a specific one. According to the authors, this approach was applied to real data from North Carolina State University, obtaining highly promising results.

Moreover, Ding et al., (2019) proposed an automated transductive transfer learning approach that addresses the poor prediction performance of models trained over data from one course and transferred to another. Their approach consists of two alternative transfer methods based on representation learning with auto-encoders: a passive approach using transductive principal component analysis and an active approach that uses a correlation alignment loss term. This work investigated the transferability of dropout prediction across similar and dissimilar courses. Finally, the obtained results were compared to known methods.

Again, in Tsiakmaki et al., (2020), the authors presented an investigation on the effectiveness of transfer learning from deep neural networks for the task of students' performance prediction in higher education. The construction of the transfer model was based on data from five compulsory courses of two undergraduate programs. Based on the experimental results, the prognosis of students at risk of failure can be achieved with satisfactory accuracy in most cases, provided that datasets of students who have attended other related courses are available.

On the other hand, López-Zambrano et al., (2020) presented a robust analysis on transfer models obtained over Moodle's data logs of 24 university courses. According to the authors, the proposed method checks whether grouping similar courses by the degree or the similar level of usage of activities provided by Moodle logs affects the portability of the prediction models. To obtain different DT models, they applied a well-known classification algorithm (i.e., Decision Tree - J48) overall datasets of the courses. In addition, they tested their portability to the other courses by comparing the obtained accuracy and loss of accuracy evaluation measures (i.e., the AUC and the loss of AUC, being the difference between two AUC values). Their analysis has shown that the portability of models is feasible under some circumstances, such as: acceptable accuracy, and more importantly, the transfer occurs between courses of the same degree.

In a recent work (López-Zambrano et al., 2021), the authors proposed the use of ontologies to improve the performance of transferable ML models. The proposal applied high-level attributes with more semantic meaning and taxonomy of actions that summarises students' interactions with Moodle. The proposed approach was compared against the previous results from the authors' work. In terms of accuracy, the results indicated that the later work could be applied to different courses with similar characteristics without performance loss.

Finally, it is important to emphasise that the construction of the dataset from courses data logs to be used to train different models is vital, as all related work reported here has pointed out this aspect. Furthermore, the interpretability that can be extracted from the model is equally important, as in the case of López-Zambrano et al., (2020), they were able to identify the attributes used to construct the resulting tree. Although this last work has answered some issues, others still remain unanswered. Regarding interpretability, other classification algorithms can also be analysed; regarding evaluation measures, other more harmonic measures can be used; and last, regarding model training and testing phases, a hold-out methodology with different percentages could also contribute to a more complex analysis of portability of models. Based on these points, this work will extend in three aspects, being classifiers, evaluation measures and different percentages of data used for training and testing the models.

4. Experimental Methodology

As mentioned earlier, the main contribution of this work is to perform a robust analysis divided in three parts: 1) the ideal dataset for training a generic model to be transferred among similar courses; 2) the selection of interpretable ML classifiers; and 3) the application of harmonic performance measures for choosing the best models. In doing this analysis, we believe to be able to propose an entire pipeline from the beginning (i.e., dataset) until the application of the generic models to their similar courses. The following subsections will present the experimental methodology aspects traditionally related in the literature as important settings for Machine Learning experiments (Drummond, 2006), such as datasets, feature selection aspects, and methods.

4.1 Data Analysis and Feature Selection

Firstly, due to significant changes in Moodle's structure of the database, we decided to use data from the years 2018 and 2019. Moreover, as the ULADECH university[1] offers courses by semesters, the data of both years were divided in semesters (i.e., 2018.01, 2018.02, 2019.01 and 2019.02).

For the selection of undergraduate courses, we considered the total amount of activities and resources created for each one of them. Initially, we took into account eight types of activities and resources, such as: assign, forum, quiz, lesson, chat, wiki, pages and URL. However, three types (i.e., lesson, chat and wiki) showed very low usage rates, being discarded.

Table 1 shows the degree courses of the ULADECH university, along with the number of semesters (SEMS), the number of courses for each degree course, the number of each activity and resource related to each degree course, and the total number of activities and resources for each degree course. Based on this table, only four undergraduate courses (i.e., in blue) were selected for composing the initial dataset. Moreover, three of them were chosen based on the number of activities and resources, and the last one was due to its representation in major areas of expertise.

[1] Catholic University Los Angeles of Chimbote in Peru

Table 1. Description of degree courses, courses, activities and resources extracted from Moodle.

Activities and Resources									
#	**Degree Courses**	**SEMs**	**Courses**	**Assign**	**Forum**	**Quiz**	**Pages**	**Url**	**Total**
1	BUSINESS ADMINISTRATION	10	77	680	690	249	685	381	2,685
2	ACCOUNTING	10	77	824	845	75	838	438	3,020
3	CHILDHOOD EDUCATION	10	77	386	406	191	405	190	1,578
4	PRIMARY SCHOOL EDUCATION	10	76	13	13	9	13	5	53
5	LAW	12	74	694	701	221	698	476	2,790
6	CIVIL ENGINEERING	10	73	503	508	42	507	214	1,774
7	PHARMACY AND BIOCHEMISTRY	10	68	351	356	55	352	240	1,354
8	SYSTEMS ENGINEERING	10	65	264	270	93	253	170	1,050
9	DENTISTRY	10	65	352	359	36	355	236	1,338
10	NURSING	10	63	375	417	93	376	216	1,477
11	PSYCHOLOGY	10	63	374	383	118	378	187	1,440
12	OBSTETRICS	10	61	272	286	46	275	188	1,067

Source: Authors.

Regarding the courses, we analysed a much greater group of them, considering the semesters that the courses were offered and which degree course they belong to, and the number of activities and resources used by each of them. After this analysis, we reduced the final number to only 35 courses, taking in consideration the different academic semesters they were offered and the degree courses.

Finally, the selected courses were grouped into two groups, being 1) degree courses group divided into Business Administration, Accounting, Law and System Engineering groups (i.e., courses belonging to any one of the four chosen degree courses); and 2) activities group divided into seven, six and five (i.e., any course with 7, 6 or 5 activities and resources).

Table 2 presents 23 numeric attributes that were selected to compose the dataset of each course. These attributes belong to eight different components. The last attribute (class attribute) has two labels, being Pass or Fail. On top of that, after this phase (feature selection), we separated the entire dataset into 35 datasets, being one for each course.

Table 2. Description of the attributes.

#	Attribute	Component	#	Attribute	Component
1	Assign submit	Assignment	13	Forum view discussion	Forum
2	Assign submit for grading	Assignment	14	Forum module view	Forum
3	Assign view	Assignment	15	Page module view	Page
4	Course enroll	Course	16	Quiz attempt	Quiz
5	Course user report	Course	17	Quiz close attempt	Quiz
6	Course view	Course	18	Quiz continue attempt	Quiz
7	Folder view	Folder	19	Quiz review	Quiz
8	Forum add discussion	Forum	20	Quiz module view	Quiz
9	Forum add post	Forum	21	Quiz view summary	Quiz
10	Forum search	Forum	22	Resource module view	Resource
11	Forum subscribe created	Forum	23	Url module view	URL
12	Forum unsubscribe	Forum	24	Class	------

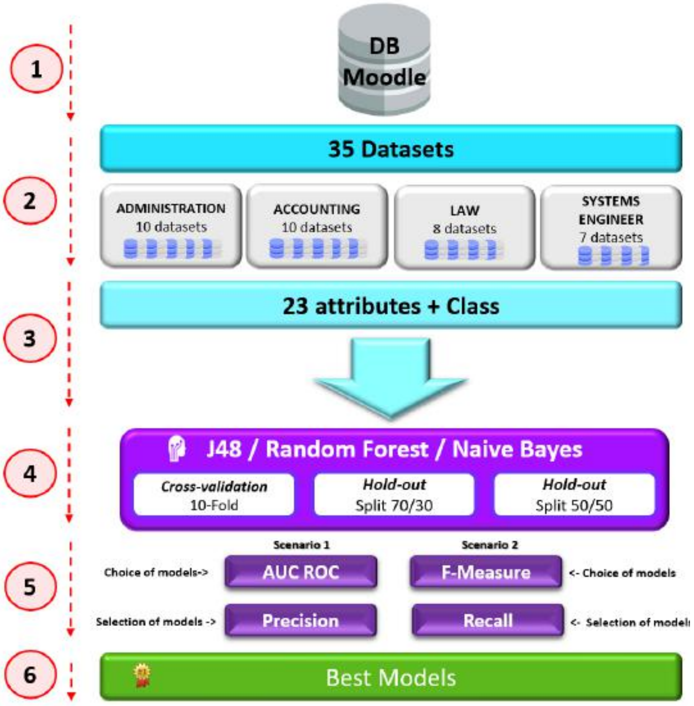
Source: Authors.

4.2 Methods and Materials

As mentioned earlier, we decided to use classification algorithms with good interpretation, such as: Decision Tree (J48), Naive Bayes and Random Forest. In this sense, the Weka ML platform was used for creating the three models with default hyper-parameters values. Then, these models were applied over the 35 datasets (i.e., one dataset for each selected course).

Two performance metrics (i.e., AUC ROC and F-Measure) were used to evaluate the models. Then, the best two models were selected from each group of courses according to those metrics (i.e., average values). Finally, Recall and Precision metrics were used to evaluate the portability performance of these two models over each group of courses, excluding the courses used to create the models. Figure 2 describes the experimental methodology used in our work.

Figure 2. Experimental methodology.



Source: Authors.

Three training and testing methodologies were applied in order to better evaluate the models (i.e., 10-Fold cross-validation, 70/30 split and 50/50 split). Usually, only a k-fold type of cross-validation is applied. However, a deeper investigation needs to consider different proportions of data to be used in training and testing. Therefore, we used these two types of hold-out methods.

5. Experimental Results

This section presents the experimental results comparing the predictive performance of the best models' portability to other courses. The results are discussed in terms of two scenarios:

- **Scenario 1:** the selection of the best models will be made by applying the AUC ROC metric for evaluating the models, and the portability performance validation of the models will be made with the Precision measure;
- **Scenario 2:** the selection of the best models will be made by applying the F-Measure metric for evaluating the models, and the portability performance validation of the models will be made with the Recall measure.

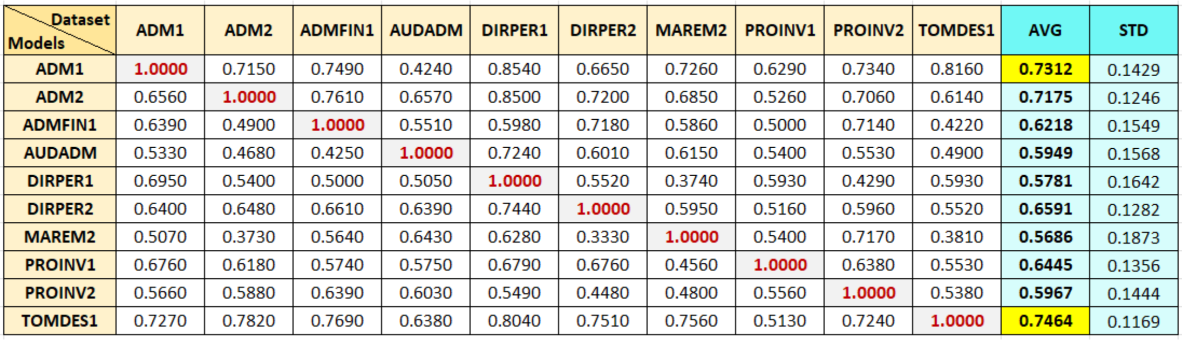
5.1 Scenario 1

The experimental results regarding scenario 1 will be discussed here in two parts. The first one is related to the selection of the best models, whereas the second one is related to the validation of the best models through the application of the Precision measure.

Table 3 presents the AUC ROC average values for J48 (decision tree) models. Note that each course represents one model for a business administration degree course. Moreover, the last two columns show the average and standard deviation of the AUC ROC values for all models of that degree course. Finally, cells in yellow show the best two models selected for further validation. In this case, our method selected the models ADM1 and TOMDES1 for business administration degree.

Table 3. The AUC ROC values for J48 models related to the Business Administration course using a 10-Fold cross-validation methodology. Columns from 2 to 11 refer to courses names (i.e. Administration 1, Administration 2, Financial Administration 1, Internal Audit, People Management 1, People Management 2, Business Marketing 2, Investment Project 1, Investment Project 2, Business Strategies 1), last two columns refer to the Average and Standard Deviation.

BUSINESS ADMINISTRATION - degree courses group



Dataset Models	ADM1	ADM2	ADMFIN1	AUDADM	DIRPER1	DIRPER2	MAREM2	PROINV1	PROINV2	TOMDES1	AVG	STD
ADM1	1.0000	0.7150	0.7490	0.4240	0.8540	0.6650	0.7260	0.6290	0.7340	0.8160	0.7312	0.1429
ADM2	0.6560	1.0000	0.7610	0.6570	0.8500	0.7200	0.6850	0.5260	0.7060	0.6140	0.7175	0.1246
ADMFIN1	0.6390	0.4900	1.0000	0.5510	0.5980	0.7180	0.5860	0.5000	0.7140	0.4220	0.6218	0.1549
AUDADM	0.5330	0.4680	0.4250	1.0000	0.7240	0.6010	0.6150	0.5400	0.5530	0.4900	0.5949	0.1568
DIRPER1	0.6950	0.5400	0.5000	0.5050	1.0000	0.5520	0.3740	0.5930	0.4290	0.5930	0.5781	0.1642
DIRPER2	0.6400	0.6480	0.6610	0.6390	0.7440	1.0000	0.5950	0.5160	0.5960	0.5520	0.6591	0.1282
MAREM2	0.5070	0.3730	0.5640	0.6430	0.6280	0.3330	1.0000	0.5400	0.7170	0.3810	0.5686	0.1873
PROINV1	0.6760	0.6180	0.5740	0.5750	0.6790	0.6760	0.4560	1.0000	0.6380	0.5530	0.6445	0.1356
PROINV2	0.5660	0.5880	0.6390	0.6030	0.5490	0.4480	0.4800	0.5560	1.0000	0.5380	0.5967	0.1444
TOMDES1	0.7270	0.7820	0.7690	0.6380	0.8040	0.7510	0.7560	0.5130	0.7240	1.0000	0.7464	0.1169

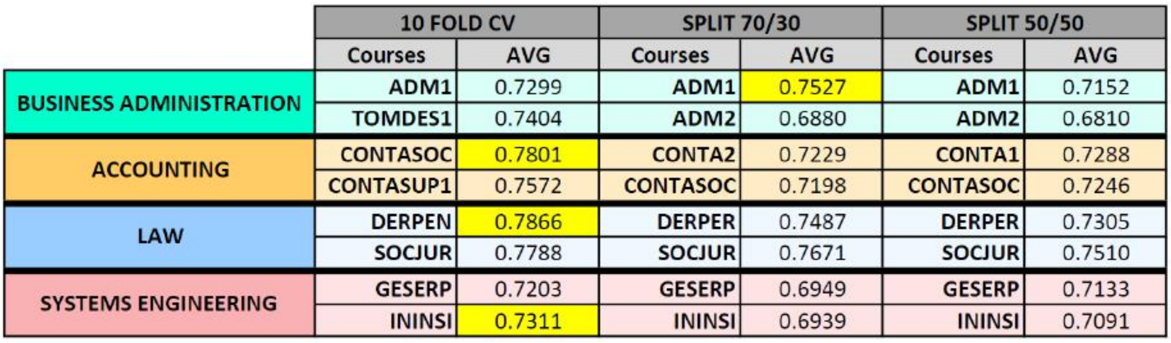
Source: Authors.

Regarding the other three degree courses, the same analysis was performed in order to select the best Decision Tree models. Table 4 shows twenty-four (24) models according to course degree and training and testing methodology. In this way, our method selected six (6) models for business administration degrees considering 10-fold CV, 70/30 and 50/50 split methodologies. The same analysis was performed in order to select six more models for accounting degrees and so forth. Moreover, cells in yellow show the best models for each degree course and for each training and testing methodology according to the AUC ROC average values.

Note that each degree course had six models selected by our method. However, in most of the cases, they are the same ones considering training and testing methodology. Moreover, in the case of business administration degree, our method selected three models (i.e., TOMDES1, ADM1 and ADM2) but ADM1 and ADM2 were selected twice in 70/30 and 50/50 split, and ADM1 was selected once in 10-fold CV, meaning that these models are indeed the most suitable for the portability to other business administration courses. Finally, note that there is a consistency in selecting the same two or three models for the other degree courses.

Table 4. The AUC ROC values for J48 models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group selection



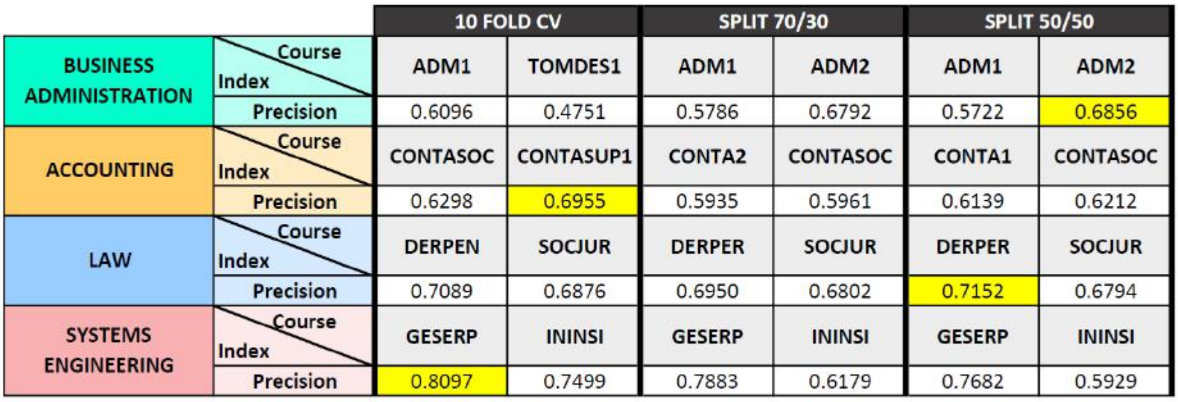
	10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
	Courses	AVG	Courses	AVG	Courses	AVG
BUSINESS ADMINISTRATION	ADM1	0.7299	ADM1	0.7527	ADM1	0.7152
	TOMDES1	0.7404	ADM2	0.6880	ADM2	0.6810
ACCOUNTING	CONTASOC	0.7801	CONTA2	0.7229	CONTA1	0.7288
	CONTASUP1	0.7572	CONTASOC	0.7198	CONTASOC	0.7246
LAW	DERPEN	0.7866	DERPER	0.7487	DERPER	0.7305
	SOCJUR	0.7788	SOCJUR	0.7671	SOCJUR	0.7510
SYSTEMS ENGINEERING	GESERP	0.7203	GESERP	0.6949	GESERP	0.7133
	ININSI	0.7311	ININSI	0.6939	ININSI	0.7091

Source: Authors.

Table 5 shows the portability performance validation results obtained when the Precision measure was applied over a degree course dataset combined by all courses belonging to that degree course excluding those used to generate the models. Note that cells in yellow indicate the best models' performance for each degree course.

Table 5. Precision values for J48 models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group portability performance validation



		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
BUSINESS ADMINISTRATION	Course Index	ADM1	TOMDES1	ADM1	ADM2	ADM1	ADM2
	Precision	0.6096	0.4751	0.5786	0.6792	0.5722	0.6856
ACCOUNTING	Course Index	CONTASOC	CONTASUP1	CONTA2	CONTASOC	CONTA1	CONTASOC
	Precision	0.6298	0.6955	0.5935	0.5961	0.6139	0.6212
LAW	Course Index	DERPEN	SOCJUR	DERPER	SOCJUR	DERPER	SOCJUR
	Precision	0.7089	0.6876	0.6950	0.6802	0.7152	0.6794
SYSTEMS ENGINEERING	Course Index	GESERP	ININSI	GESERP	ININSI	GESERP	ININSI
	Precision	0.8097	0.7499	0.7883	0.6179	0.7682	0.5929

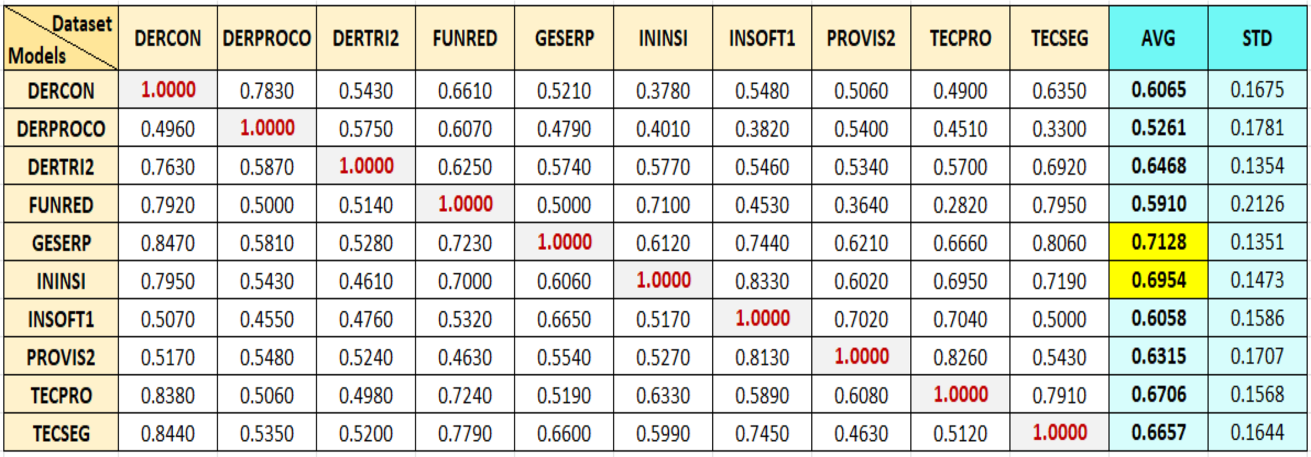
Source: Authors.

Regarding the portability performance of the models for each degree course, according to Table 5, DERPEN achieved performance above 70% for Law (0.7152), and GESERP achieved performance around 80% for Systems Engineering (0.8097). In both cases the achieved performance prediction is adequate for model portability according to some works (López-Zambrano et al., 2020; Tsiakmaki et al., 2020). However, both CONTASUP1 for Accounting (0.6955) and ADM2 for Business Administration (0.6856) achieved prediction performance below 70%.

The same analysis was performed for the activities groups (i.e., any course with 7, 6 or 5 activities and resources). Table 6 presents the AUC ROC average values for J48 (decision tree) models. Note that each course represents one model according to the number of activities. Moreover, the last two columns show the average and standard deviation of the AUC ROC values for all models. Finally, cells in yellow show the best two models. In this case, our method selected GESERP and ININSI models.

Table 6. The AUC ROC values for models related to courses with seven activities using a 10-Fold cross-validation methodology. Columns from 2 to 11 refer to courses names (i.e., Constitutional Law, Constitutional Procedural Law, Tax Law 2, Networking Fundamentals, ERP Management, Introduction to Systems Engineering, Software Engineering 1, Visual programming 2, Programming Techniques, Technology and Network Security), last two columns refer to Average and Standard Deviation.

Activity group 07 - Courses with seven activities



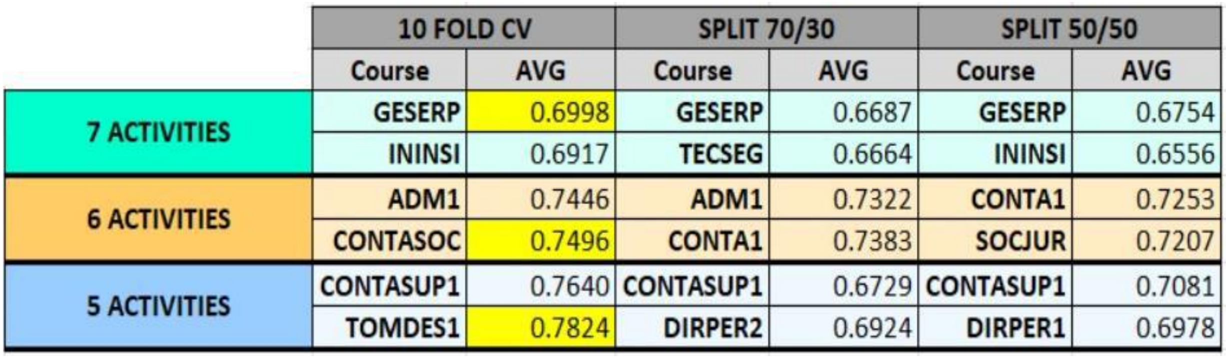
Dataset Models	DERCON	DERPROCO	DERTRI2	FUNRED	GESERP	ININSI	INSOFT1	PROVIS2	TECPRO	TECSEG	AVG	STD
DERCON	1.0000	0.7830	0.5430	0.6610	0.5210	0.3780	0.5480	0.5060	0.4900	0.6350	**0.6065**	0.1675
DERPROCO	0.4960	1.0000	0.5750	0.6070	0.4790	0.4010	0.3820	0.5400	0.4510	0.3300	**0.5261**	0.1781
DERTRI2	0.7630	0.5870	1.0000	0.6250	0.5740	0.5770	0.5460	0.5340	0.5700	0.6920	**0.6468**	0.1354
FUNRED	0.7920	0.5000	0.5140	1.0000	0.5000	0.7100	0.4530	0.3640	0.2820	0.7950	**0.5910**	0.2126
GESERP	0.8470	0.5810	0.5280	0.7230	1.0000	0.6120	0.7440	0.6210	0.6660	0.8060	**0.7128**	0.1351
ININSI	0.7950	0.5430	0.4610	0.7000	0.6060	1.0000	0.8330	0.6020	0.6950	0.7190	**0.6954**	0.1473
INSOFT1	0.5070	0.4550	0.4760	0.5320	0.6650	0.5170	1.0000	0.7020	0.7040	0.5000	**0.6058**	0.1586
PROVIS2	0.5170	0.5480	0.5240	0.4630	0.5540	0.5270	0.8130	1.0000	0.8260	0.5430	**0.6315**	0.1707
TECPRO	0.8380	0.5060	0.4980	0.7240	0.5190	0.6330	0.5890	0.6080	1.0000	0.7910	**0.6706**	0.1568
TECSEG	0.8440	0.5350	0.5200	0.7790	0.6600	0.5990	0.7450	0.4630	0.5120	1.0000	**0.6657**	0.1644

Source: Authors.

Table 7 shows eighteen (18) models according to activity group and training and testing methodology. In this way, our method selected six (6) models for seven activity groups considering 10-fold CV, 70/30 and 50/50 split methodologies. The same analysis was performed in order to select six more models for six activity groups and so forth. Moreover, cells in yellow show the best model for each group.

Table 7. The AUC ROC values for J48 models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups selection



	10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
	Course	AVG	Course	AVG	Course	AVG
7 ACTIVITIES	GESERP	0.6998	GESERP	0.6687	GESERP	0.6754
	ININSI	0.6917	TECSEG	0.6664	ININSI	0.6556
6 ACTIVITIES	ADM1	0.7446	ADM1	0.7322	CONTA1	0.7253
	CONTASOC	0.7496	CONTA1	0.7383	SOCJUR	0.7207
5 ACTIVITIES	CONTASUP1	0.7640	CONTASUP1	0.6729	CONTASUP1	0.7081
	TOMDES1	0.7824	DIRPER2	0.6924	DIRPER1	0.6978

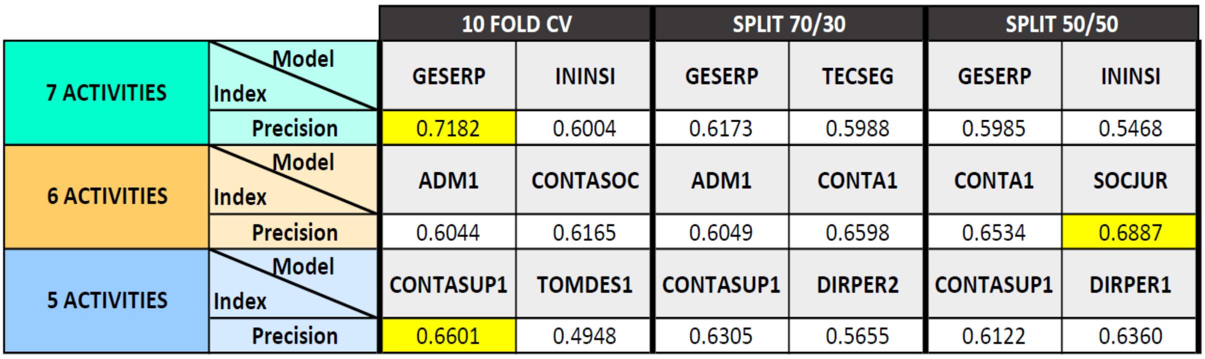
Source: authors. Source: Authors.

Note that each activity group had six models selected by our method. However, in most of the cases, they are the same ones considering training and testing methodologies. Moreover, in the case of group of seven activities, our method selected three models (i.e. GESERP, ININSI and TECSEG), but GESERP was selected three times in 10-fold CV, 70/30 and 50/50 split, meaning that this model is indeed the most suitable for the portability to other courses in the same group.

Regarding the portability performance of the models for each activity group, according to Table 8, GESERP achieved performance above 70% for seven activity groups (0.7182). However, both SOCJUR for the group of six activities (0.6887) and CONTASUP1 for the group of five activities (0.6601) achieved prediction performance below 70%. On top of that, considering the predictive results showed on both tables (5 and 8), we can point out that the GESERP model achieved the best results for both degree course group (0.8097) and the group of seven activities (0.7182). This might indicate that the number of activities offered in Moodle plays an important role for building robust models to be transferred to the same group of courses or activities.

Table 8. Precision values for J48 models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups portability performance validation



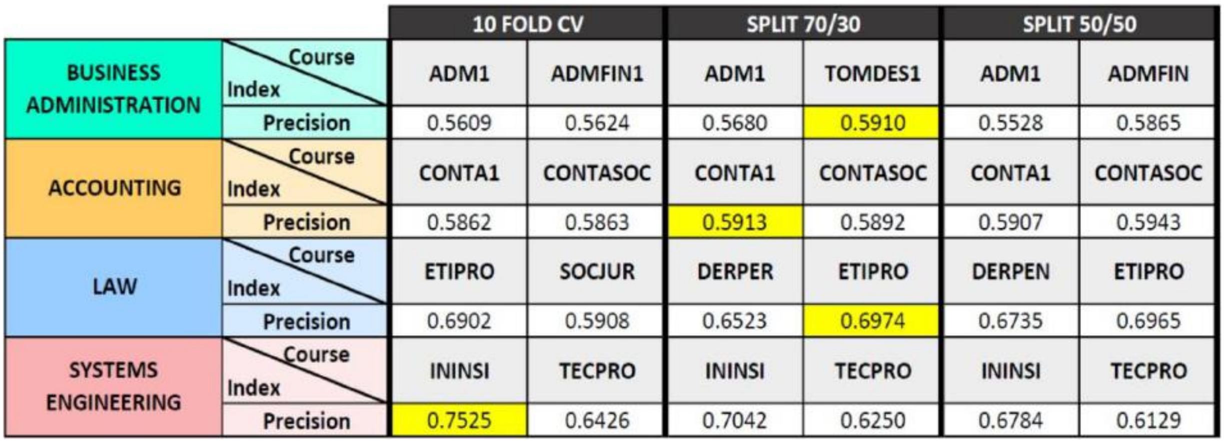
		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
7 ACTIVITIES	Model Index	GESERP	ININSI	GESERP	TECSEG	GESERP	ININSI
	Precision	0.7182	0.6004	0.6173	0.5988	0.5985	0.5468
6 ACTIVITIES	Model Index	ADM1	CONTASOC	ADM1	CONTA1	CONTA1	SOCJUR
	Precision	0.6044	0.6165	0.6049	0.6598	0.6534	0.6887
5 ACTIVITIES	Model Index	CONTASUP1	TOMDES1	CONTASUP1	DIRPER2	CONTASUP1	DIRPER1
	Precision	0.6601	0.4948	0.6305	0.5655	0.6122	0.6360

Source: Authors.

Analysing the portability performance of models created with Naive Bayes (NB) and Random Forest (RF) classification techniques, we can point out the occurrence of similar results achieved by models created with J48 and shown on previous tables. Firstly, considering only NB models, Tables 9 and 10 show that the ININSI model achieved the best results for both degree course group (0.7525) and group of seven activities (0.7350). Secondly, the other models (i.e., presented in both tables) have achieved predictive performance below 70%.

Table 9. The Precision values for Naive Bayes (NB) models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group portability performance validation

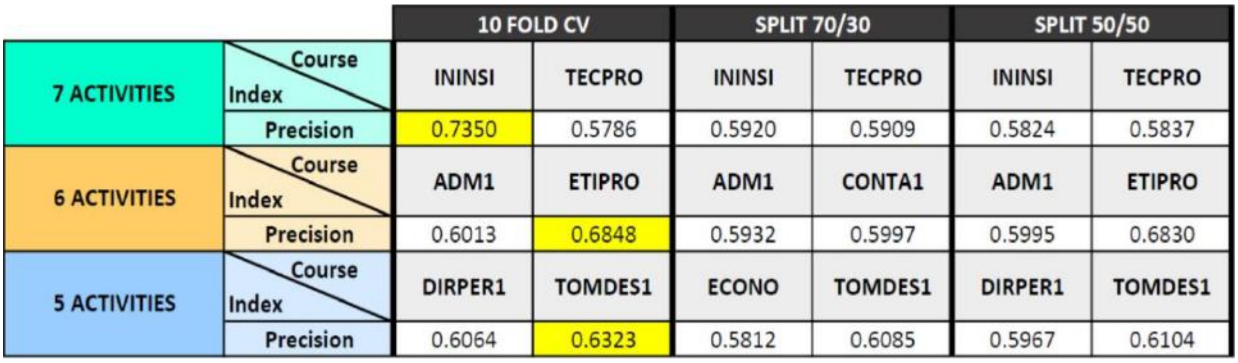


		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
BUSINESS ADMINISTRATION	Course Index	ADM1	ADMFIN1	ADM1	TOMDES1	ADM1	ADMFIN
	Precision	0.5609	0.5624	0.5680	0.5910	0.5528	0.5865
ACCOUNTING	Course Index	CONTA1	CONTASOC	CONTA1	CONTASOC	CONTA1	CONTASOC
	Precision	0.5862	0.5863	0.5913	0.5892	0.5907	0.5943
LAW	Course Index	ETIPRO	SOCJUR	DERPER	ETIPRO	DERPEN	ETIPRO
	Precision	0.6902	0.5908	0.6523	0.6974	0.6735	0.6965
SYSTEMS ENGINEERING	Course Index	ININSI	TECPRO	ININSI	TECPRO	ININSI	TECPRO
	Precision	0.7525	0.6426	0.7042	0.6250	0.6784	0.6129

Source: Authors.

Table 10. The Precision values for Naive Bayes (NB) models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups portability performance validation



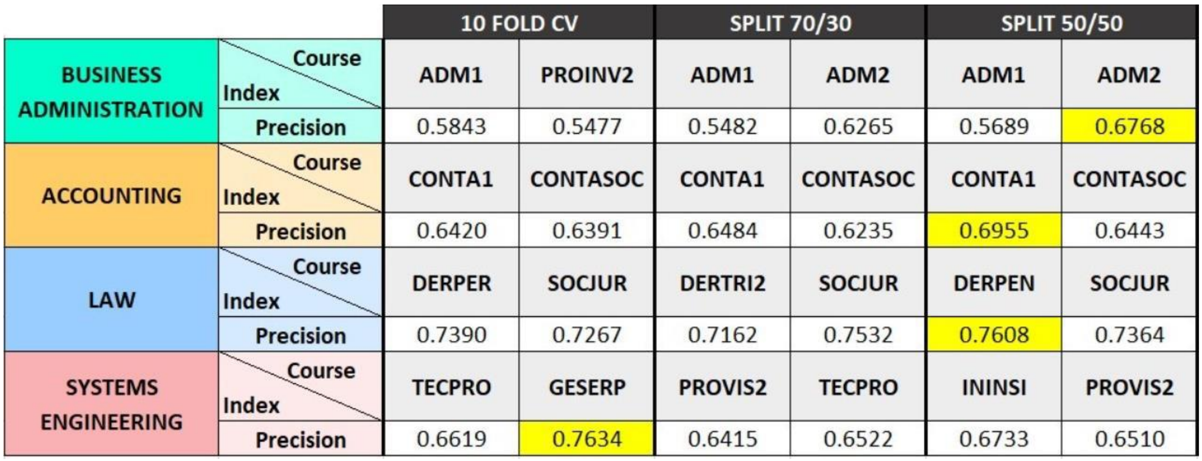
		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
7 ACTIVITIES	Course Index	ININSI	TECPRO	ININSI	TECPRO	ININSI	TECPRO
	Precision	0.7350	0.5786	0.5920	0.5909	0.5824	0.5837
6 ACTIVITIES	Course Index	ADM1	ETIPRO	ADM1	CONTA1	ADM1	ETIPRO
	Precision	0.6013	0.6848	0.5932	0.5997	0.5995	0.6830
5 ACTIVITIES	Course Index	DIRPER1	TOMDES1	ECONO	TOMDES1	DIRPER1	TOMDES1
	Precision	0.6064	0.6323	0.5812	0.6085	0.5967	0.6104

Source: Authors.

Regarding the performance RF models, Table 11 shows that DERPER for Law (0.7634) and GESERP for Systems Engineering(0.7390) achieved performance above 70%. However, both CONTA1 for Accounting (0.6955) and ADM2 for Business Administration (0.6768) achieved prediction performance below 70%. Moreover, Table 12 shows that all three models for the activities group achieved predictive performance below 70%, being TECSEG for the group of seven activities (0.6308), CONTA1 for the group of six activities (0.6856), and DIRPER1 for the group of five activities (0.6685).

Table 11. The Precision values for Random Forest (RF) models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group validation

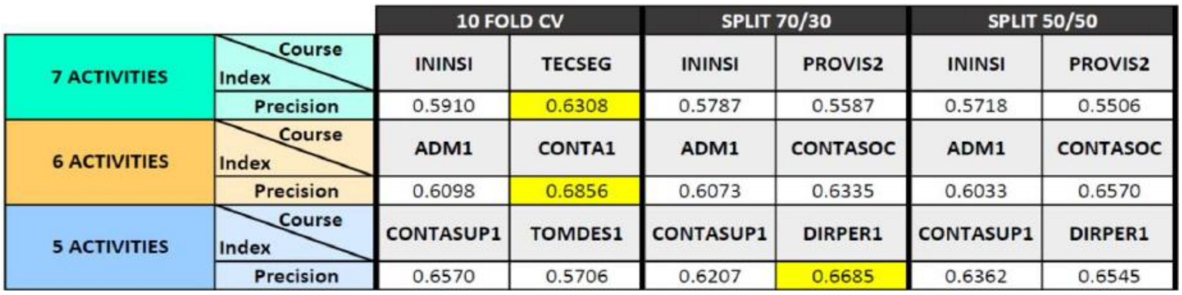


		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
BUSINESS ADMINISTRATION	Course Index	ADM1	PROINV2	ADM1	ADM2	ADM1	ADM2
	Precision	0.5843	0.5477	0.5482	0.6265	0.5689	0.6768
ACCOUNTING	Course Index	CONTA1	CONTASOC	CONTA1	CONTASOC	CONTA1	CONTASOC
	Precision	0.6420	0.6391	0.6484	0.6235	0.6955	0.6443
LAW	Course Index	DERPER	SOCJUR	DERTRI2	SOCJUR	DERPEN	SOCJUR
	Precision	0.7390	0.7267	0.7162	0.7532	0.7608	0.7364
SYSTEMS ENGINEERING	Course Index	TECPRO	GESERP	PROVIS2	TECPRO	ININSI	PROVIS2
	Precision	0.6619	0.7634	0.6415	0.6522	0.6733	0.6510

Source: Authors.

Table 12. The Precision values for Random Forest (RF) models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups validation



		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
7 ACTIVITIES	Course / Index	ININSI	TECSEG	ININSI	PROVIS2	ININSI	PROVIS2
	Precision	0.5910	0.6308	0.5787	0.5587	0.5718	0.5506
6 ACTIVITIES	Course / Index	ADM1	CONTA1	ADM1	CONTASOC	ADM1	CONTASOC
	Precision	0.6098	0.6856	0.6073	0.6335	0.6033	0.6570
5 ACTIVITIES	Course / Index	CONTASUP1	TOMDES1	CONTASUP1	DIRPER1	CONTASUP1	DIRPER1
	Precision	0.6570	0.5706	0.6207	0.6685	0.6362	0.6545

Source: Authors.

Finally, considering all models created by J48, NB and RF classification techniques and being validated by the Precision measure, we can point out that Law and Systems Engineering courses obtained the best overall results (i.e., DERPER and ETIPRO from Law, and GESERP and ININSI from Systems).

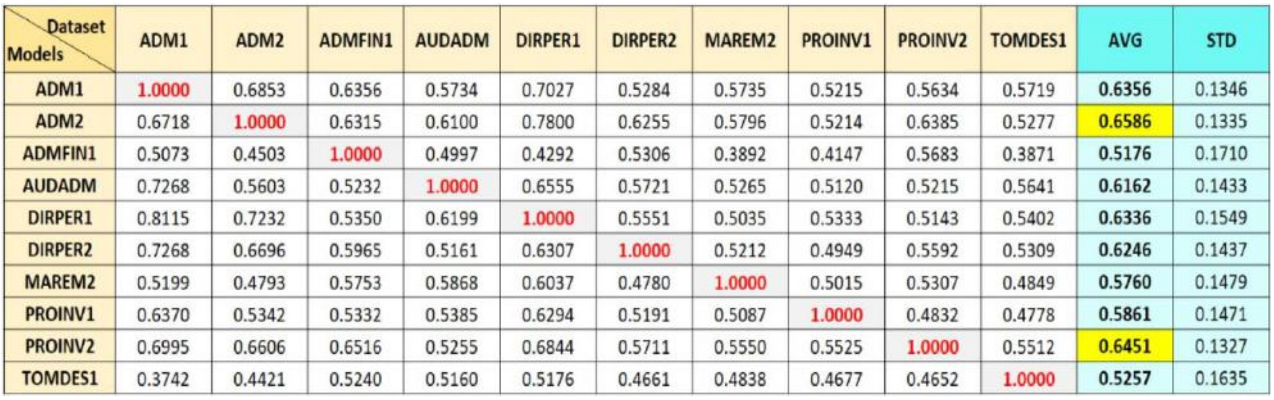
From the activities group's perspective, we can state that models with seven activities obtained the best results in two out of three cases, being GESERP created with J48 and ININSI created with NB. Although none of the RF models reported on Table 12 achieved results higher than 68%, the CONTA1 model from group of six activities obtained the best overall result among all activities models created by the RF technique.

5.2 Scenario 2

Again the experimental results will be discussed here in two parts. The first one is related to the selection of the best models, whereas the second one is related to the portability performance validation of the best models through the application of the Precision measure. Table 13 presents the F-measure average values for J48 models.

Table 13. The F-measure values for J48 models related to Business Administration course using a 10-Fold cross-validation methodology. Columns from 2 to 11 refer to courses names (i.e., Administration 1, Administration 2, Financial Administration 1, Internal Audit, People Management 1, People Management 2, Business Marketing 2, Investment Project 1, Investment Project 2, Business Strategies 1), last two columns refer to Average and Standard Deviation.

BUSINESS ADMINISTRATION - degree courses group



Dataset / Models	ADM1	ADM2	ADMFIN1	AUDADM	DIRPER1	DIRPER2	MAREM2	PROINV1	PROINV2	TOMDES1	AVG	STD
ADM1	1.0000	0.6853	0.6356	0.5734	0.7027	0.5284	0.5735	0.5215	0.5634	0.5719	**0.6356**	0.1346
ADM2	0.6718	1.0000	0.6315	0.6100	0.7800	0.6255	0.5796	0.5214	0.6385	0.5277	**0.6586**	0.1335
ADMFIN1	0.5073	0.4503	1.0000	0.4997	0.4292	0.5306	0.3892	0.4147	0.5683	0.3871	**0.5176**	0.1710
AUDADM	0.7268	0.5603	0.5232	1.0000	0.6555	0.5721	0.5265	0.5120	0.5215	0.5641	**0.6162**	0.1433
DIRPER1	0.8115	0.7232	0.5350	0.6199	1.0000	0.5551	0.5035	0.5333	0.5143	0.5402	**0.6336**	0.1549
DIRPER2	0.7268	0.6696	0.5965	0.5161	0.6307	1.0000	0.5212	0.4949	0.5592	0.5309	**0.6246**	0.1437
MAREM2	0.5199	0.4793	0.5753	0.5868	0.6037	0.4780	1.0000	0.5015	0.5307	0.4849	**0.5760**	0.1479
PROINV1	0.6370	0.5342	0.5332	0.5385	0.6294	0.5191	0.5087	1.0000	0.4832	0.4778	**0.5861**	0.1471
PROINV2	0.6995	0.6606	0.6516	0.5255	0.6844	0.5711	0.5550	0.5525	1.0000	0.5512	**0.6451**	0.1327
TOMDES1	0.3742	0.4421	0.5240	0.5160	0.5176	0.4661	0.4838	0.4677	0.4652	1.0000	**0.5257**	0.1635

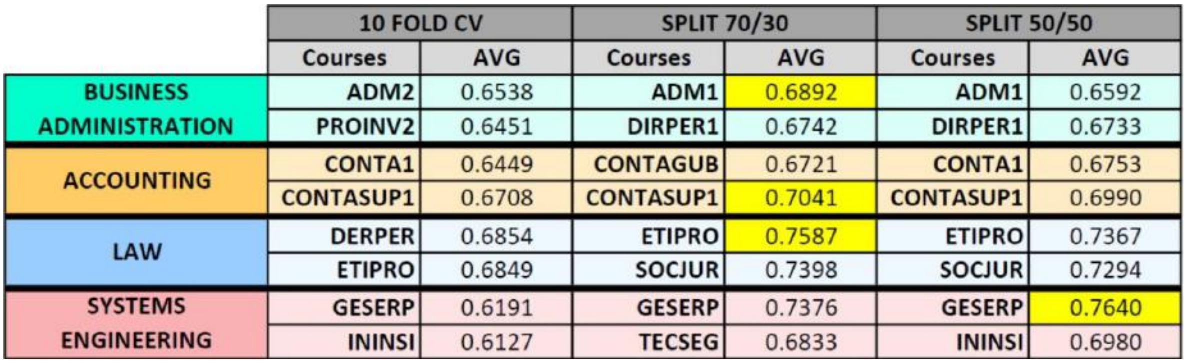
Source: Authors.

Note that each course represents one model for a Business Administration degree course. Moreover, the last two columns show the average and standard deviation of the F-measure values for all models of that degree course. Finally, cells in yellow show the best two models selected for further validation. In this case, our method selected the models ADM2 and PROINV2.

Again the same analysis was performed in order to select the best Decision Tree (J48) models for the other three degree courses. Table 14 shows twenty-four (24) models according to course degree and training and testing methodology. In this way, our method selected six (6) models for Business Administration degree considering 10-fold CV, 70/30 and 50/50 split methodologies. The same selection was performed in order to select six more models for Accounting degree and so forth. Moreover, cells in yellow show the best models for each degree course and for each training and testing methodology according to the F-measure average values.

Table 14. The F-measure values for J48 models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group selection



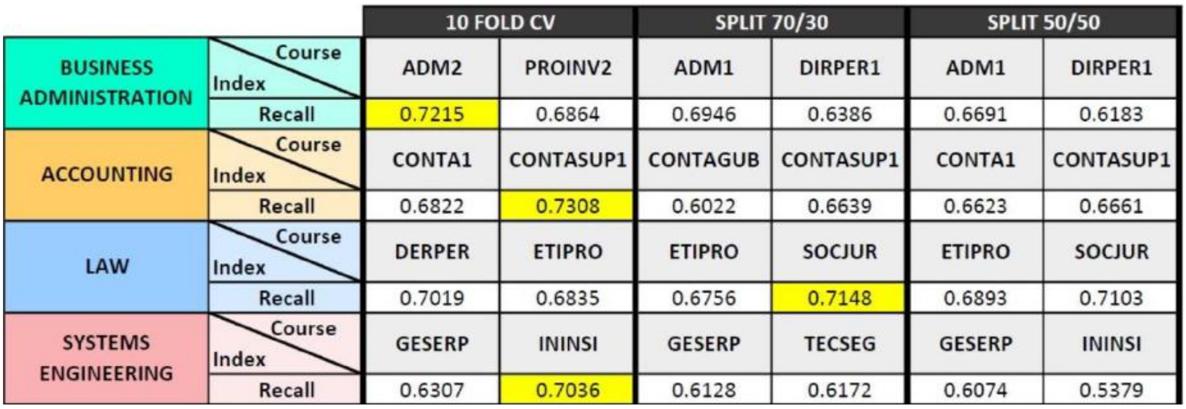
	10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
	Courses	AVG	Courses	AVG	Courses	AVG
BUSINESS ADMINISTRATION	ADM2	0.6538	ADM1	0.6892	ADM1	0.6592
	PROINV2	0.6451	DIRPER1	0.6742	DIRPER1	0.6733
ACCOUNTING	CONTA1	0.6449	CONTAGUB	0.6721	CONTA1	0.6753
	CONTASUP1	0.6708	CONTASUP1	0.7041	CONTASUP1	0.6990
LAW	DERPER	0.6854	ETIPRO	0.7587	ETIPRO	0.7367
	ETIPRO	0.6849	SOCJUR	0.7398	SOCJUR	0.7294
SYSTEMS ENGINEERING	GESERP	0.6191	GESERP	0.7376	GESERP	0.7640
	ININSI	0.6127	TECSEG	0.6833	ININSI	0.6980

Source: Authors.

Table 15 shows the validation results obtained when the Recall measure was applied over a degree course dataset combined by all courses belonging to that degree course excluding those used to generate the models. Note that cells in yellow indicate the best models' performance for each degree course.

Table 15. The Recall values for J48 models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group portability performance validation



		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
BUSINESS ADMINISTRATION	Course Index	ADM2	PROINV2	ADM1	DIRPER1	ADM1	DIRPER1
	Recall	0.7215	0.6864	0.6946	0.6386	0.6691	0.6183
ACCOUNTING	Course Index	CONTA1	CONTASUP1	CONTAGUB	CONTASUP1	CONTA1	CONTASUP1
	Recall	0.6822	0.7308	0.6022	0.6639	0.6623	0.6661
LAW	Course Index	DERPER	ETIPRO	ETIPRO	SOCJUR	ETIPRO	SOCJUR
	Recall	0.7019	0.6835	0.6756	0.7148	0.6893	0.7103
SYSTEMS ENGINEERING	Course Index	GESERP	ININSI	GESERP	TECSEG	GESERP	ININSI
	Recall	0.6307	0.7036	0.6128	0.6172	0.6074	0.5379

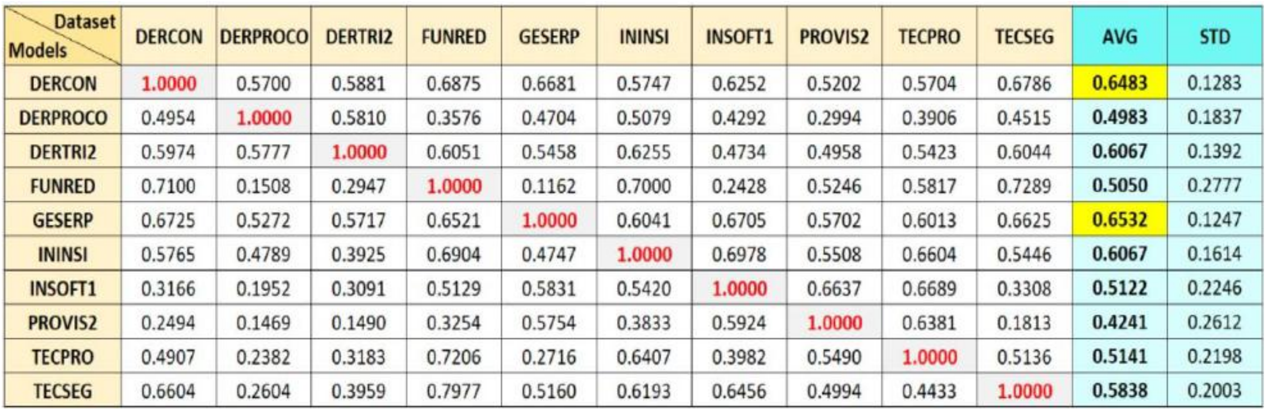
Source: Authors.

Regarding the portability performance of the models for each degree course (i.e., shown in Table 15), all four selected models achieved performance above 70%, being ADM1 from Business Administration (0.7215), CONTASUP1 for Accounting (0.7308), SOCJUR for Law (0.7148), and ININSI for Systems Engineering (0.7036).

Again the same analysis was performed for the activities groups (i.e., any course with 7, 6 or 5 activities and resources). Table 16 presents the F-measure average values for J48 (decision tree) models. As already discussed, each course represents one model according to the number of activities. Moreover, the last two columns show the average and standard deviation of the F-measure values for all models. Finally, cells in yellow show the best two models. In this case, our method selected DERCON and GESERP models.

Table 16. The F-measure values for models related to courses with seven activities using a 10-Fold cross-validation methodology. Columns from 2 to 11 refer to courses names (i.e., Constitutional Law, Constitutional Procedural Law, Tax Law 2, Networking Fundamentals, ERP Management, Introduction to Systems Engineering, Software Engineering 1, Visual programming 2, Programming Techniques, Technology and Network Security), last two columns refer to Average and Standard Deviation.

Activity group 07 - Courses with seven activities



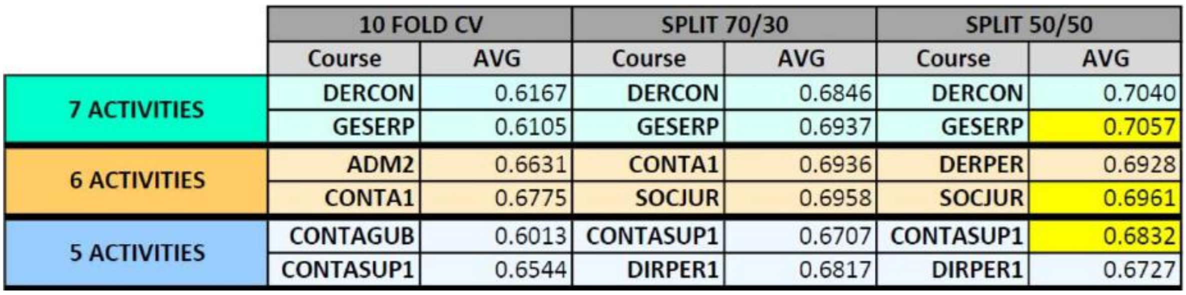
Dataset Models	DERCON	DERPROCO	DERTRI2	FUNRED	GESERP	ININSI	INSOFT1	PROVIS2	TECPRO	TECSEG	AVG	STD
DERCON	1.0000	0.5700	0.5881	0.6875	0.6681	0.5747	0.6252	0.5202	0.5704	0.6786	**0.6483**	0.1283
DERPROCO	0.4954	1.0000	0.5810	0.3576	0.4704	0.5079	0.4292	0.2994	0.3906	0.4515	**0.4983**	0.1837
DERTRI2	0.5974	0.5777	1.0000	0.6051	0.5458	0.6255	0.4734	0.4958	0.5423	0.6044	**0.6067**	0.1392
FUNRED	0.7100	0.1508	0.2947	1.0000	0.1162	0.7000	0.2428	0.5246	0.5817	0.7289	**0.5050**	0.2777
GESERP	0.6725	0.5272	0.5717	0.6521	1.0000	0.6041	0.6705	0.5702	0.6013	0.6625	**0.6532**	0.1247
ININSI	0.5765	0.4789	0.3925	0.6904	0.4747	1.0000	0.6978	0.5508	0.6604	0.5446	**0.6067**	0.1614
INSOFT1	0.3166	0.1952	0.3091	0.5129	0.5831	0.5420	1.0000	0.6637	0.6689	0.3308	**0.5122**	0.2246
PROVIS2	0.2494	0.1469	0.1490	0.3254	0.5754	0.3833	0.5924	1.0000	0.6381	0.1813	**0.4241**	0.2612
TECPRO	0.4907	0.2382	0.3183	0.7206	0.2716	0.6407	0.3982	0.5490	1.0000	0.5136	**0.5141**	0.2198
TECSEG	0.6604	0.2604	0.3959	0.7977	0.5160	0.6193	0.6456	0.4994	0.4433	1.0000	**0.5838**	0.2003

Source: Authors.

Table 17 shows eighteen (18) models according to activity group and training and testing methodology. In this way, our method selected six (6) models for the group of seven activities considering 10-fold CV, 70/30 and 50/50 split methodologies. The same analysis was performed in order to six more models for the group of six activities and so forth. Moreover, cells in yellow show the best model for each group.

Table 17. The F-measure values for J48 models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups selection



	10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
	Course	AVG	Course	AVG	Course	AVG
7 ACTIVITIES	DERCON	0.6167	DERCON	0.6846	DERCON	0.7040
	GESERP	0.6105	GESERP	0.6937	GESERP	0.7057
6 ACTIVITIES	ADM2	0.6631	CONTA1	0.6936	DERPER	0.6928
	CONTA1	0.6775	SOCJUR	0.6958	SOCJUR	0.6961
5 ACTIVITIES	CONTAGUB	0.6013	CONTASUP1	0.6707	CONTASUP1	0.6832
	CONTASUP1	0.6544	DIRPER1	0.6817	DIRPER1	0.6727

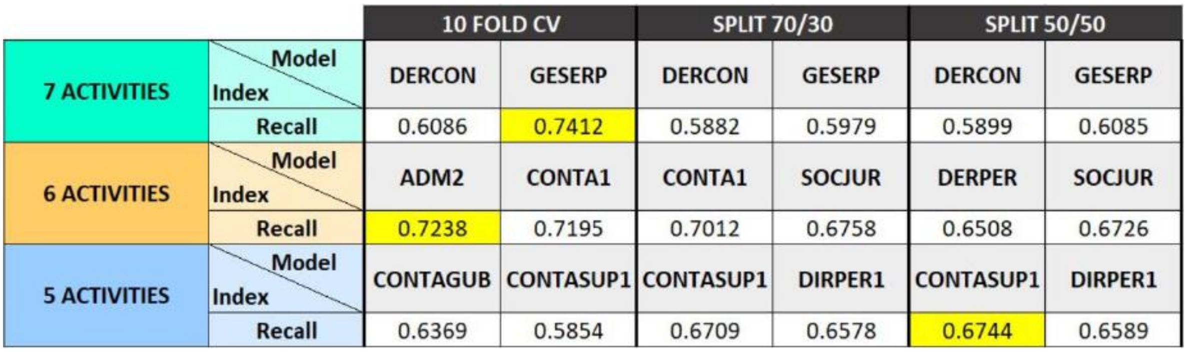
Source: Authors.

Considering the results presented in Table 18, note that each activity group had six models selected by our method. In most of the cases, they are the same ones considering training and testing methodologies. Moreover, in the case of group of seven activities our method selected only two models (i.e., DERCON and GESERP). On top of that, both methods were selected three times in 10-fold CV, 70/30 and 50/50 split, meaning that these models are indeed the most suitable for the portability to other courses in the same group.

Regarding the performance of the models for each activity group, according to Table 18, GESERP for the group of seven activities (0.7412) and ADM2 for the group of six activities (0.7238) achieved performance above 70%. However, CONTASUP1 for the group of five activities (0.6744) achieved prediction performance below 70%. Finally, we can point out that no model has achieved the best results in both degree course and group of seven activities as it occurred during the scenario one's analysis for the J48 classification technique.

Table 18. The Recall values for J48 models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups portability performance validation



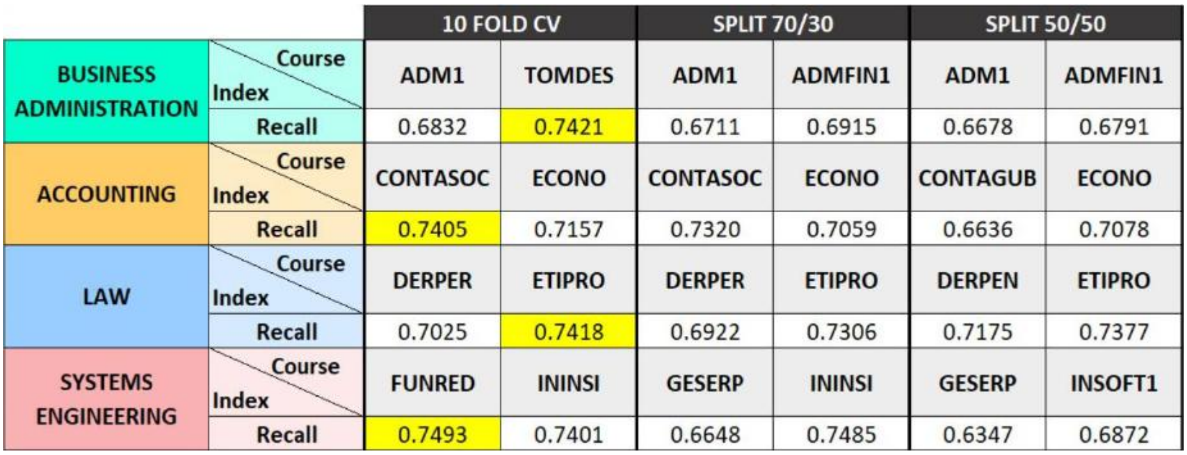
		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
7 ACTIVITIES	Model Index	DERCON	GESERP	DERCON	GESERP	DERCON	GESERP
	Recall	0.6086	0.7412	0.5882	0.5979	0.5899	0.6085
6 ACTIVITIES	Model Index	ADM2	CONTA1	CONTA1	SOCJUR	DERPER	SOCJUR
	Recall	0.7238	0.7195	0.7012	0.6758	0.6508	0.6726
5 ACTIVITIES	Model Index	CONTAGUB	CONTASUP1	CONTASUP1	DIRPER1	CONTASUP1	DIRPER1
	Recall	0.6369	0.5854	0.6709	0.6578	0.6744	0.6589

Source: Authors.

Analysing the portability performance of models created with Naive Bayes (NB) and Random Forest (RF) classification techniques, we can point out there has been no similarity compared to scenario one. However, regarding the performance of NB models, Tables 19 and 20 show that all models have achieved performance results above 70%. Such performance has not occurred in scenario 1.

Table 19. The Recall values for Naive Bayes (NB) models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group portability performance validation

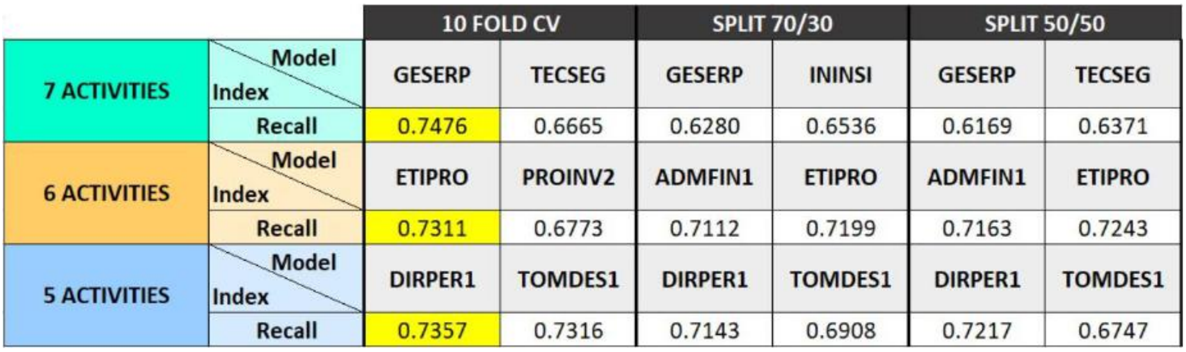


		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
BUSINESS ADMINISTRATION	Index \ Course	ADM1	TOMDES	ADM1	ADMFIN1	ADM1	ADMFIN1
	Recall	0.6832	0.7421	0.6711	0.6915	0.6678	0.6791
ACCOUNTING	Index \ Course	CONTASOC	ECONO	CONTASOC	ECONO	CONTAGUB	ECONO
	Recall	0.7405	0.7157	0.7320	0.7059	0.6636	0.7078
LAW	Index \ Course	DERPER	ETIPRO	DERPER	ETIPRO	DERPEN	ETIPRO
	Recall	0.7025	0.7418	0.6922	0.7306	0.7175	0.7377
SYSTEMS ENGINEERING	Index \ Course	FUNRED	ININSI	GESERP	ININSI	GESERP	INSOFT1
	Recall	0.7493	0.7401	0.6648	0.7485	0.6347	0.6872

Source: Authors.

Table 20. The Recall values for Naive Bayes (NB) models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups portability performance validation



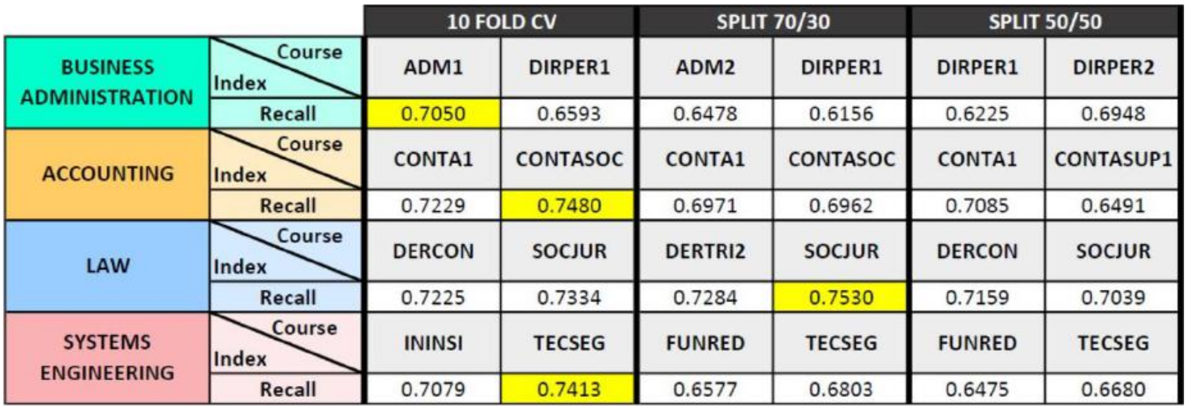
		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
7 ACTIVITIES	Index \ Model	GESERP	TECSEG	GESERP	ININSI	GESERP	TECSEG
	Recall	0.7476	0.6665	0.6280	0.6536	0.6169	0.6371
6 ACTIVITIES	Index \ Model	ETIPRO	PROINV2	ADMFIN1	ETIPRO	ADMFIN1	ETIPRO
	Recall	0.7311	0.6773	0.7112	0.7199	0.7163	0.7243
5 ACTIVITIES	Index \ Model	DIRPER1	TOMDES1	DIRPER1	TOMDES1	DIRPER1	TOMDES1
	Recall	0.7357	0.7316	0.7143	0.6908	0.7217	0.6747

Source: Authors.

Regarding the performance of RF models, Tables 21 and 22 show that all models have achieved performance results above 70%. Such performance has not occurred in scenario 1, considering all three classification techniques evaluated in this analysis.

Table 21. The Recall values for Random Forest (RF) models related to all four degree courses using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Degree courses group portability performance validation

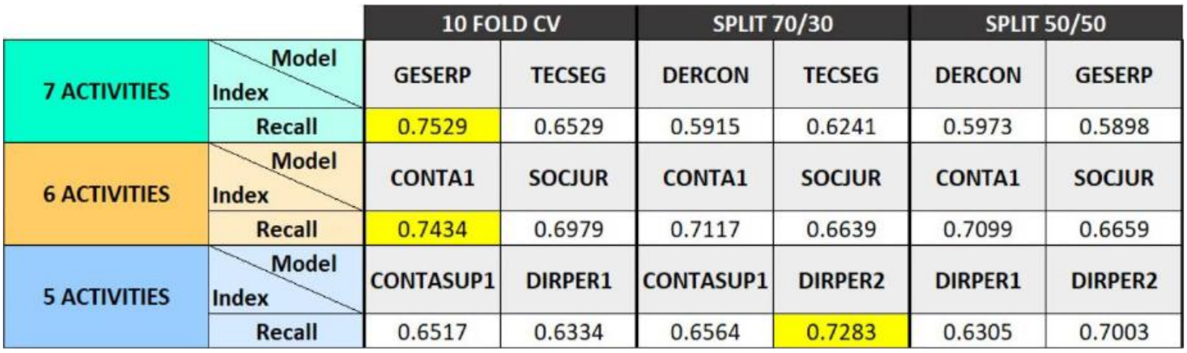


		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
BUSINESS ADMINISTRATION	Course / Index	ADM1	DIRPER1	ADM2	DIRPER1	DIRPER1	DIRPER2
	Recall	0.7050	0.6593	0.6478	0.6156	0.6225	0.6948
ACCOUNTING	Course / Index	CONTA1	CONTASOC	CONTA1	CONTASOC	CONTA1	CONTASUP1
	Recall	0.7229	0.7480	0.6971	0.6962	0.7085	0.6491
LAW	Course / Index	DERCON	SOCJUR	DERTRI2	SOCJUR	DERCON	SOCJUR
	Recall	0.7225	0.7334	0.7284	0.7530	0.7159	0.7039
SYSTEMS ENGINEERING	Course / Index	ININSI	TECSEG	FUNRED	TECSEG	FUNRED	TECSEG
	Recall	0.7079	0.7413	0.6577	0.6803	0.6475	0.6680

Source: Authors.

Table 22. The Recall values for Random Forest (RF) models related to all three activity groups using a 10-Fold cross-validation, 70/30 split and 50/50 split.

Activity groups portability performance validation



		10 FOLD CV		SPLIT 70/30		SPLIT 50/50	
7 ACTIVITIES	Model / Index	GESERP	TECSEG	DERCON	TECSEG	DERCON	GESERP
	Recall	0.7529	0.6529	0.5915	0.6241	0.5973	0.5898
6 ACTIVITIES	Model / Index	CONTA1	SOCJUR	CONTA1	SOCJUR	CONTA1	SOCJUR
	Recall	0.7434	0.6979	0.7117	0.6639	0.7099	0.6659
5 ACTIVITIES	Model / Index	CONTASUP1	DIRPER1	CONTASUP1	DIRPER2	DIRPER1	DIRPER2
	Recall	0.6517	0.6334	0.6564	0.7283	0.6305	0.7003

Source: Authors.

5.3 Portability Performance Results

We now analyse the relative frequency with which different models (J48, Nb and RF) using AUC ROC and F-measure metrics have been selected by our method in both scenarios. Starting with degree courses group, ADM1 models for Business Administration degree course were selected 15 times out of 18, CONTASOC models for Accounting were selected 13 times out of 18, DERPEN models for Law were selected 11 times out of 18, and ININSI models for Systems Engineering were selected 12 times out 18.

The same relative frequency analysis was performed for the activities group. Starting with the group of seven activities, GESERP models were selected 11 times out of 18, CONTA1 models for six activities were selected 9 times out of 18, and CONTASUP models for five activities were selected 11 times out of 18.

Overall, considering the performance results shown in tables of both scenarios, the GESERP model created with Decision Tree (J48) classification technique, selected with AUC ROC metric and validated with Precision measure, obtained the best performance result (0.8097) for Systems Engineering degree course. Moreover, this model (the same configuration) has also achieved the best result for the group of seven activities (i.e., greater number of activities).

This finding raises an important question concerning its performance using different configurations (i.e., classification technique, selection and validation measures). Regarding AUC ROC for selection and Precision for validation, the GESERP model created with RF classification technique has obtained again the best performance result for Systems Engineering degree course, meaning three (3) cases in six (6) possible cases.

Analysing scenario 2 (i.e., F-measure for selection, Recall for validation and all three classification techniques), GESERP models have obtained the best performance results again in three (3) cases out of six (6) possible cases. Moreover, it is important to emphasise that the predictive performance of GESERP models (J48, NB and RF) were the best in all three group of seven activities (i.e., one for each classification technique).

Finally, the GESERP models have obtained robust results, indicating that they are suitable for transfer predictive models to Systems Engineering degree course group when the AUC ROC metric is used for selection, and the Precision measure is used for validation. On top of that, GESERP models are suitable for portability to group of seven activities when the F-measure metric is used for selection, and the Recall measure is used for validation.

6. Conclusion and Future Work

The main goal of this work was to apply a Transfer Learning approach on Learning Management Systems logs (i.e., Moodle) in order to achieve good portability of models, and then be able to predict the performance of undergraduate students. Two different scenarios have been implemented considering selection and validation phases of the models. The AUC ROC and Precision predictive accuracy measures were used as selection and validation methods in scenario 1, whereas F-measure and Recall were used as selection and validation methods in scenario 2.

An empirical analysis was conducted in order to evaluate the performance of the models created with three well-known classification algorithms (i.e., Decision Tree, Random Forest and Naive Bayes) over different datasets representing the courses of the same degree and the activities offered by each course. In this sense, we can evaluate the effectiveness of applying the portability of models by using different interpretable Machine Learning techniques over Moodle data logs from distinct university courses.

In experiments using 35 classification datasets representing courses belonging to four different degree courses, overall, ADM1 models for Business Administration degree course (15 out of 18), CONTASOC models for Accounting (13 out of 18), DERPEN models for Law were (11 out of 18), and ININSI models for Systems Engineering (12 out 18) were the most selected models.

When considering the activities group and the courses related to each group, GESERP models for seven activities (11 out of 18), CONTA1 models for six activities (9 out of 18), and CONTASUP models for five activities were the most selected models.

Overall, considering the portability performance results of both scenarios, the GESERP model created with Decision Tree (J48) classification technique, selected with AUC ROC metric and validated with Precision measure obtained the best performance result (0.8097). Regarding the best models of each group, based on the performance results and the relative frequency with which they have been selected, we can state that they can be used for portability of models.

As a direction for future work, it would be interesting to carry out experiments with more datasets representing courses of other degree courses. In this way, our evaluation could be extended as we used only four degree courses.

References

Ahmed, D. M., Abdulazeez, A. M., Zeebaree, D. Q., & Ahmed, F. Y. (2021, June). Predicting University's Students Performance Based on Machine Learning Techniques. In *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)* (pp. 276-281). IEEE.

Aleksandrova, Y. (2019). Predicting Students Performance in Moodle Platforms Using Machine Learning Algorithms. In *Conferences of the department Informatics* (No. 1, pp. 177-187). Publishing house Science and Economics Varna.

Ayala, J. C., & Manzano, G. (2018). Academic performance of first-year university students: The influence of resilience and engagement. *Higher Education Research & Development*, *37*(7), 1321-1335.

Barros, R. C., De Carvalho, A. C., & Freitas, A. A. (2015). *Automatic design of decision-tree induction algorithms*. Springer.

Boyer, S., & Veeramachaneni, K. (2015, June). Transfer learning for predictive models in massive open online courses. In *International conference on artificial intelligence in education* (pp. 54-63). Springer, Cham.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Büchner, A. (2016). Moodle Administration. An administrator's guide of configuring, securing, customizing and extending Moodle. *Packt Publishing, Birmingham, S*, *9*, 41.

Davidson, R. A. (2002). Relationship of study approach and exam performance. *Journal of Accounting Education*, *20*(1), 29-44.

Ding, M., Wang, Y., Hemberg, E., & O'Reilly, U. M. (2019, March). Transfer learning using representation learning in massive open online courses. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 145-154).

Dougiamas, M., & Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system. In *EdMedia+ innovate learning* (pp. 171-178). Association for the Advancement of Computing in Education (AACE).

Drummond, C. (2006, July). Machine learning as an experimental science (revisited). In AAAI workshop on evaluation methods for machine learning (pp. 1-5).

Fernández-Berrocal, P., & Checa, P. (2016). Emotional intelligence and cognitive abilities. *Frontiers in psychology*, *7*, 955.

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. *The journal of machine learning research*, *15*(1), 3133-3181.

Hao, J., Gan, J., & Zhu, L. (2022). MOOC performance prediction and personal performance improvement via Bayesian network. *Education and Information Technologies*, 1-24.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

Hunt, X. J., Kabul, I. K., & Silva, J. (2017, August). Transfer learning for education data. In *Proceedings of the ACM SIGKDD Conference, El Halifax, NS, Canada* (Vol. 17).

López-Zambrano, J., Lara, J. A., & Romero, C. (2020). Towards portability of models for predicting students' final performance in university courses starting from moodle logs. *Applied Sciences*, *10*(1), 354.

López-Zambrano, J., Lara, J. A., & Romero, C. (2021). Improving the portability of predicting students' performance models by using ontologies. *Journal of Computing in Higher Education*, 1-19.

Mangini, C. G., Lima, N. D. da S., & Nääs, I. de A. . (2021). Thermal mapp routing in pharmaceutical products transportation using machine learning approach: a systematic review. *Research, Society and Development*, *10*(16), e170101623665. https://doi.org/10.33448/rsd-v10i16.23665

Mascarenhas, T. A. T. ., Moriel Junior , J. G. ., Gomes, R. de S. R. ., & Mello, G. J. (2020). Application of machine learning algorithms in the Classification of Specialized Knowledge of Physics Teachers. *Research, Society and Development*, *9*(11), e86191110584. https://doi.org/10.33448/rsd-v9i11.10584

Neves, A. R. N. das, Okada, H. K. R., & Shitsuka, R. (2019). Gesture Recognition in Images Using Neural Networks. *Research, Society and Development*, *8*(11), e278111470. https://doi.org/10.33448/rsd-v8i11.1470

Nguyen, V. A., Nguyen, Q. B., & Nguyen, V. T. (2018, August). A model to forecast learning outcomes for students in blended learning courses based on learning analytics. In *Proceedings of the 2nd International Conference on E-Society, E-Education and E-Technology* (pp. 35-41).

Ossani, P. C., Rossoni, D. F. ., Cirillo, M. Ângelo ., & Borém, F. M. . (2021). Classification of specialty coffees using machine learning techniques . *Research, Society and Development*, *10*(5), e13110514732. https://doi.org/10.33448/rsd-v10i5.14732

Palma, L. C., de Oliveira, L. M., & Viacava, K. R. (2011). Sustainability in Brazilian federal universities. *International Journal of Sustainability in Higher Education*.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345-1359.

Piña, A. A. (2012). An overview of learning management systems. *Virtual Learning Environments: Concepts, methodologies, tools and applications*, 33-51.

Quinn, R. J., & Gray, G. (2020). Prediction of student academic performance using Moodle data from a Further Education setting. *Irish Journal of Technology Enhanced Learning*, *5*(1).

Raga Jr, R. C., & Raga, J. D. (2017). Monitoring Class Activity and Predicting Student Performance Using Moodle Action Log Data. *International Journal of Computing Sciences Research*, *1*(3), 1-16.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368-384.

Sarkar, D., Bali, R., & Ghosh, T. (2018). *Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras*. Packt Publishing Ltd.

Teles, W. de S. ., Machado, A. P. ., Cantos Júnior, P. C. C. ., Melo, C. M. de ., Silva, M. H. S. ., Silva, R. N. da ., & Jeraldo, V. de L. S. . (2021). Machine learning and automatic selection of attributes for the identification of Chagas disease from clinical and sociodemographic data. *Research, Society and Development*, *10*(4), e19310413879. https://doi.org/10.33448/rsd-v10i4.13879

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. *Applied Sciences*, *10*(6), 2145.

Ülker, D., & Yılmaz, Y. (2016). Learning Management Systems and Comparison of Open Source Learning Management Systems and Proprietary Learning Management Systems. *Journal of Systems Integration (1804-2724)*, *7*(2).

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, *3*(1), 1-40.