Kufa Journal of Engineering
Vol. 16, No. 3, July 2025, P.P. 292-311
Article history: Received 18 September 2024, last revised 23 November 2024, accepted 23 November 2024

PREDICT LEARNERS' PERFORMANCE USING AN ONTOLOGICAL-BASED MODEL ON AN E-LEARNING PLATFORM

Safa Ridha Albo Abdullah

Department of Cyber Security, College of Information Technology, University of Babylon, Babil, Iraq. Email: safaruda@uobabylon.edu.iq.

https://doi.org/10.30572/2018/KJE/160316

ABSTRACT

In learning analytics and educational data mining, a prominent challenge is posed by the lack of portability and transferability of predictive models across different courses. A novel ontology-based decision tree model is introduced in this study, which significantly enhances portability by incorporating semantic features. Unlike conventional decision tree models that are static and course-specific, the approach allows for the dynamic generation and adaptation of decision tree rules using ontologies, thus enabling seamless application across multiple courses without compromising accuracy. The core innovation of the method lies in the use of the Semantic Web Rule Language (SWRL) to integrate decision tree rules into an ontological framework, allowing reasoning and adaptation based on domain-specific semantics. By embedding decision tree rules that differentiate between pass and fail outcomes within an ontology, superior portability, and adaptability are achieved across courses with similar usage patterns. Applied to the Open University Learning Design (OULD) science courses dataset, advanced machine learning techniques were utilized to derive predictive rules, achieving an exceptional prediction accuracy of 97%. This demonstrates not only the effectiveness of the model but also its potential for robust transferability across various educational contexts, marking a significant advancement over existing methods.

KEYWORDS

Student Performance, Transfer Learning, Model Portability, Ontology, SWRL, Machine Learning, Decision Tree.

1. INTRODUCTION

Online learning has been well-known as substitute to traditional face-to-face learning, so it represents a combination of modern education and information technologies. Moreover, it is regarded essential to the advancement of educational equity (Hussain et al., 2018). Online learning is extensively employed worldwide due to its high degree of flexibility in terms of time and place, low bar for knowledge acquisition, and abundance of learning resources (Qiu et al., 2022). Its platforms enable students conducting a variety of learning activities, including reading, downloading, submitting papers, uploading content, and designing and delivering presentations (Chweya et al., 2020). Without being required to attend a class, students can access learning materials via online learning systems (Hamim, Benabbou and Sael, 2021). Nevertheless, this does not mean that online learning is without its restrictions. The lack of direct interaction has been indicated as a factor of academic failure or low grades among some students (Gheisari et al., 2023). Furthermore, students' dropout rates are higher in online learning courses than in traditional settings (Ullah et al., 2023).

Consequently, previous literature invites proposing efficient approaches to predict learners' performance at an early stage (Aslam et al., 2021). An early prediction is a recently recognized phenomena that involves anticipating learners' outcomes to assist learners avoiding incompletion of their online courses. However, an accurate prediction relies on understanding features that may affect their achievement. According to the literature (Sultana, Rani and Farquad, 2019), predicting students' success is a problematical task because of the huge amounts of data in educational and learning management databases. As such, the application of Semantic Web may open the door for better prediction performance (Fadhil, Abed and Jasim, 2021).

An extension of the standard Web is Semantic Web, functioning as an indispensable tool for increasing and strengthening the comprehension of information between people and computers (Sultana, Rani and Farquad, 2019). Educational systems employ Semantic Web technologies to improve the value of educational content and provide learning activities such as ontologies and semantic rules that are tailored to the needs of each student (Bolock, Abdennadher and Herbert, 2021). In the education context, ontologies have been developed to gather data and categorise learning content, thereby facilitating human-machine communication (Zeebaree et al., 2019). Thus, there are numerous advantages to be gained by executing ontologies in online or e-learning, such as enhancing students' retention, implementing timely interventions to assist students at risk of failure, defining attributes that may affect students' academic performance, and improving the quality of education in practice (Iqbal et al., 2024). In another study

(Tkachenko et al., 2024), ontologies were proposed to create conceptual maps to expose students to different experiences.

This present research aims to evaluate the portability degree of a model built based on ontologies. The proposed model is constructed from students' interactions with Moodle logs. The idea of portability is the capacity to apply information from a single environment directly to another. Within the educational sphere, this concept is extremely useful since it allows applying a model discovered on a prior course (source) to another new course (target) (Buddhika, Karunarathne and Nanayakkara, 2024). Most earlier efforts on a model portability have used a Transfer Learning (TL) technique, which involves a tune-up procedure, so that the updated model is transferred from one course to another. Other works employ a Generalization strategy, which seeks to establish a single generic model that fits all previously completed courses (Mohammed, Kareem and Mohammed, 2022). This is why this research used the terms "portability" or "transferability" instead of the related term "generalization", since it is believed that this represents the straightforward application of a model gained with one dataset to a different dataset.

The development of predictive models for educational data mining has made significant strides, but a persistent challenge is the lack of model portability across different courses (Aslam et al., 2021). Conventional models are often tailored to specific courses, requiring considerable re-training when applied to new contexts. This leads to inefficiency and a reduced ability to generalize across various educational environments (Aslam et al., 2021). Furthermore, existing machine learning techniques, including decision trees and support vector machines, while accurate, lack the semantic flexibility needed to adapt across diverse datasets. Predicting learners' performance based on either traditional machine learning or deep learning techniques requires training all available data to obtain a high accuracy (Zeebaree et al., 2019). This present research, however, aims to address this issue by applying the proposed ontological-based model that is developed for one course to predict the performance of another course. Therefore, a new approach in classifying students' performance is proposed based on an ontological engineering and Semantic Web Rule Language (SWRL) rules to achieve courses' portability and transferability. Based on the data gathered from the interactions of students with the Learning Management System (LMS), the predictive model attempts to forecast whether or not a student will succeed in a specific course (pass or fail) (Papadakis, 2023). This research has many contributions, in comparison to previous literature. First, it proposes a new general ontological-based model to predict students' performance of several different courses based on an ontological network built for one course only. Second, the study uses high-level attributes with more semantic meaning, such as ontologies by creating semantic rules (decision rule). Third, it generates new features because develop the predictability of the proposed model and to decrease the complexity of the methods that are used. Moreover, this research constructs optimized machine learning-based models to improve the performance accuracy. Finally, it provides an early prediction system of the performance students' in Virtual Learning Environment (VLEs) to avoid students' failure or dropout.

In Section 2 the theoretical backdrop and earlier research are described. The research methodology is explained in Section 3. Section 4 reports the results and discussion. This study concludes in Section 5 and suggests possible next directions.

2. THEORETICAL BACKGROUND AND LITERATURE REVIEW

2.1. Machine learning techniques

As predictive models, several machine learning approaches have been used in the past. Previous research (Qiu et al., 2022) indicates that machine learning may successfully and correctly forecast how well kids will do on digital platforms. The rules in this study were extracted using the decision tree (DT) approach.

2.1.1. Decision-tree (DT)

Based on directed nodes and edges, a decision tree is defined as a hierarchical construction that can manage categorical and numerical data (Abdullah and Al-Azawei, 2023). The leaf nodes of a decision tree indicate a specific class, whereas the root and other internal nodes reflect data that is classified using features and test conditions (Mohammed, Kareem and Mohammed, 2022). Several decision tree techniques are built using Hunt's algorithm, and the C5.0 algorithm was employed in this case study (Gheisari et al., 2023).

2.1.2. Random search algorithm (RS)

The random search optimizer is one instance of a Monte Carlo technique. An optimization technique called random search is utilized to detect the optimal set of hyperparameters for a given model of the machine-learning (Lee, Cheang and Moslehpour, 2022). The arbitrary search technique is applied to regression and classification problems in many domains to find the optimal hyperparameters that can improve prediction accuracy (Abdullah and Al-Azawei, 2023). This study aims to maximize the DT classifier's performance by utilizing the optimizer of random search (Fadhil, Abed and Jasim, 2021). Each model's random initialization of a collection of hyperparameters is the first step in random search

2.2. Related work

2.2.1. Machine learning approach

Various techniques have been proposed in the literature to improve the accuracy of student performance prediction. Traditional decision trees have been widely adopted due to their simplicity, but they suffer from poor portability. Machine learning models such as neural networks and SVMs can deliver high accuracy but lack semantic adaptability (Hamzah and Dhannoon, 2021). Some researchers have explored ontology-based approaches; however, these models tend to be static and do not easily generalize across different educational courses (Merchant *et al.*, 2022). Our proposed method overcomes these limitations by dynamically integrating decision tree rules within an ontological framework, allowing for greater portability and predictive accuracy across courses

A model was developed by Daud et al. (Daud *et al.*, 2017) to predict whether or not students will finish their degrees. We looked at the suggested model with the use of demographic data. Support Vector Machine (SVM) outperformed the other methods, with an F1 score of 0.867. The results of the study also indicated that the best indicators were location, natural gas usage, self-employment, and electricity spending. In order to anticipate students' results, Umer et al. (Umer et al., 2018) examined the effectiveness of four predictive models: K-Nearest Neighbor (KNN), Naive Bayes (NB), Random Forest (RF), and Linear Discriminant Analysis (LDA). Every week, the prediction was made using assignment scores and engagement data from VLE. After the first week the prediction accuracy was 70%, the data showed that the assignment scores were the strongest discriminative variable. Furthermore, the outcomes presented that Random Forest consistently beat other models every week. The efficacy of gradient-boosted trees, J48, JRIP, decision trees, CART, and Nave Bayes in forecasting students' involvement at the Open University (OU) in a social science course was examined by Hussain et al. (Hussain *et al.*, 2020). The J48 algorithm was shown to have the best accuracy, coming up at 88.52%. Soni et al. (Soni *et al.*, 2018) developed a model that used NB, DT, and SVM to assess student performance based on their most recent output. Twenty features out of the 48 features were chosen for the extraction procedure. With 83.33% accuracy, support vector machines outperformed Decision Trees and Naïve Bayes models.

2.2.2. The semantic modelling approach

To achieve multiple goals, including failure/abundance prediction, decision-making, and orientation, Hamim et al. (Hamim, Benabbou and Sael, 2021) proposed a generic ontology for modeling the student profile. They also proposed a system that associates ontology with machine learning, utilizing an algorithm based on decision trees and SWRL rules. The C5.0 algorithm was employed in this case study. The model's attained results are 81.9% in recall and 83.6% in accuracy. In another research study, to enhance the predictability and transferability of students' performance prediction models, López-Zambrano et al. (López-Zambrano, Lara and Romero, 2022) introduced an ontology-based model. Their dataset included 1,840 students from 16 different courses offered by the computer science department at Cordoba University. The results indicated that the proposed ontology significantly improved the model's prediction accuracy, achieving a rate of 62%. However, to the research findings cannot be generalized beyond engineering and computer science courses. In a research study conducted by (Abd El-Rady, 2020), the utility of an ontological model was combined with data mining methods, including decision trees, random forests, and J48 to predict students' performance. The dataset comprised students' interactions with Facebook groups and a learning management system. The study found that two approaches were used to evaluate model based on the precision-recall matrix the first approach recall 92.3% and precision 85.7%, while the second approach recall 94.5% and precision 98.8%. Finally, Grivokostopoulou et al. (Grivokostopoulou, Perikos and Hatzilygeroudis, 2014) developed an educational framework that utilizes ontologies and semantic rules to increase the quality of a curriculum and the learning activities provided to each student.

3. RESEARCH METHODOLOGY

This research aims to present a new approach in classifying students' performance based on ontological engineering and SWRL rules over two time periods. These are quarter 1 [Q1] and quarter 4 [Q4]) that were proposed for the one course dataset namely, OULA Science courses. The suggested technique uses a particular process to improve the classifiers' prediction accuracy in order to accomplish this goal. As illustrated in Fig.1, this calls for preprocessing the data, based on the features that are already available new features were created, applying an optimization technique to determine the ideal hyperparameters, using the approach of a feature selection to identify superior features that may have an impact on learners' performance, and building the ontology, SWRL rules, and semantic reasoner classifier. The efficacy of the suggested technique in improving the models' performance is demonstrated.

3.1. Dataset

As mentioned previously, the Open University (OU) dataset was used in this study (Kuzilek, Hlosta and Zdrahal, 2017). This dataset contains demographic data, course enrollment information, assessment results, clickstream data, and final performance scores for training machine learning models. The Open Data Institute (ODI) has formally verified this dataset (Kuzilek, Hlosta and Zdrahal, 2017), which is freely accessible via OULAD. Within the OULA dataset, data related to clickstream, courses, students, and registrations are distributed across seven tables, with students serving as the primary focus of the information. Over a period of nine months, data on 32,593 students were collected. Students were able to choose from seven modules, each of which was offered at least twice a year.

The seven tables containing the raw data were linked by identification columns and provided information on assessments, types of assessments, submission dates for assessments, courses, types of VLE materials, clickstreams showing students' VLE engagement, course registration details, and students' demographic information. In the current study, two classes—"pass" and "fail"—were created from the reported student performance to provide a binary categorization. The class labels for withdrawal and failure were combined into one label, "fail," while the pass and distinction labels were merged into one label, "pass," to produce this categorization. This approach was taken to prevent the issue of class imbalance. Every subject was separated into quartiles (Abdullah and Al-Azawei, 2023), and early support was provided to students who were most likely to fail.

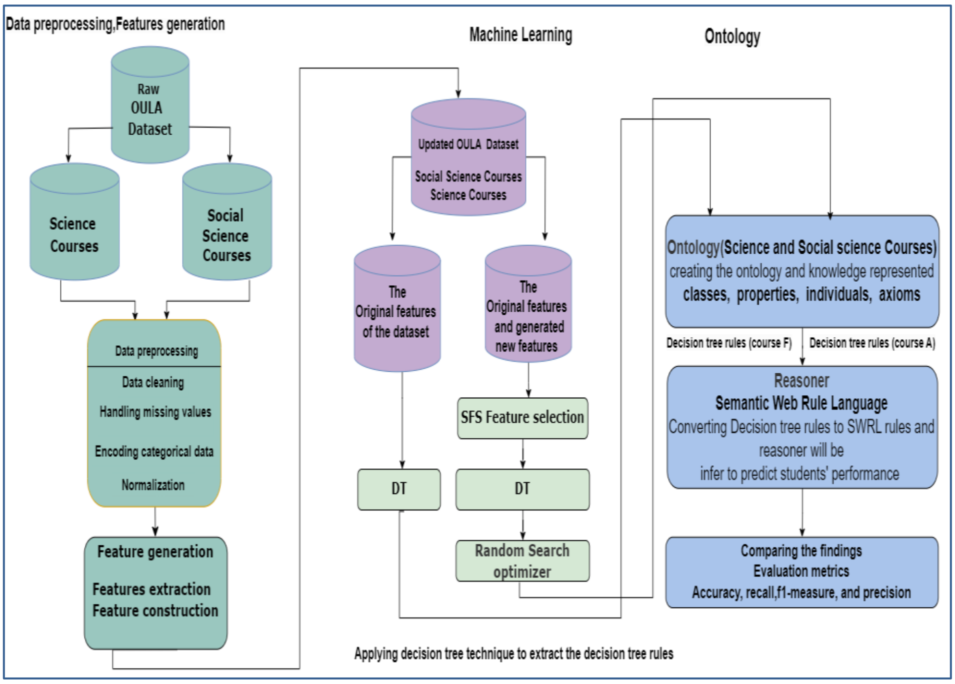

Fig.1. The block diagram of the proposed approach

3.2. Data preprocessing

The dataset was obtained from a large number of data files in an unprocessed, organized format (Abdullah and Al-Azawei, 2023). However, before the data were utilized in the machine learning algorithms, they were pre-processed. Preprocessing consists of a variety of methods and approaches , including data-cleaning, addressing missing values, categorical data-coding, and detecting inconsistent data . Data-cleaning involves removing features such as the code module, Identification (ID) number, and code presentation, which are not necessary for prediction (Qasrawi et al., 2021). Some values were missing from deprivation band (IMD band) attributes and/or the assessment scores of the dataset used. Based on the Open University

(OU's) claim of omission for all assessment values that students fail to enter, 'all –1' was entered to indicate these missing assessment values. The IMD band feature was considered as replacing any missing data (Zhou et al., 2024). Using representations of numerical that machine learning algorithms can understand, categorical variables were transformed into encoded data. This dataset included values encoded ordinally for the code module, highest education, code presentation, and age band, while gender, region, and handicap were encoded nominally. Normalization was done for every value of numeric feature that would be used as the input for the techniques of machine learning (Tan, Steinbach and Kumar, 2006). This included multiple features, such as dataplus, collaboration, forum, content, resources, 'homepage, glossary, number of previous attempts, subpages, and URLs. To ensure that all the feature values fell within a single range, this step was performed according to Eq. 1 (Abdullah and Al-Azawei, 2023).

$$\hat{V} = \frac{V - min_A}{max_A - min_A} \qquad (1)$$

The terms minA and maxA reflect the original values lowest and maximum, respectively, for every given feature, if the variable V indicates the value of the feature.

3.3. Feature generation

In order for these new predictors to efficiently collect more significant information, the generation of feature is the act of creating new attributes from original features (Zhou et al., 2024). Three of the most often used techniques for producing attribute are: 1) Data mapping to a new space; 2) Extraction of the feature; and 3) Creation of feature (Tan, Steinbach and Kumar, 2006).

3.3.1. Feature extraction

Three categories were used to group the attributes: performance, demographics, and behavior. Behavioral characteristics were extracted utilizing the students' VLE table, and interactions with each site were added to the data concerning the types of sites. The type of site was described using the VLE database, which included all ID sites and their corresponding types. For each course, 20 different types of ID sites were identified. Consequently, twenty new qualities were acquired, and data were gathered on students' interactions with these activities over two separate quarters. The course did not formally begin until Q1, followed by the remaining quarters. All ID sites identified with various kinds of activities were retained in the same column, and the ID of the interacted site was saved without mentioning its type. The extent of students' interactions with each activity type was recorded separately. Thus, the feature extraction process was utilized to assess students' interactions with various VLE activity types. The types of ID sites in the departments of social science and science courses included 'region,' 'imd_band,' 'studied_credits,' 'disability,' 'gender,' 'highest_education,' 'ageband,' 'num_of_prev_attempts,' 'Folder,' 'SharedSubPage,' 'ExternalQuiz,' 'OuContent,' 'OuCollaborate,' 'DataPlus,' 'DualPane,' 'Forumng,' 'Glossary,' 'HomePage,' 'HtmlActivity,' 'Ouelluminate,' 'OuWiki,' 'Page,' 'Questionnaire,' 'Quiz,' 'RepeatActivity,' 'Resource,' 'Subpage,' and 'Url.' Therefore, twenty new features were obtained. Students' interactions with these activities were calculated at five different intervals. The second period was designated as quarter 1 (Q1), spanning from the first day to 53 days of the course. The fourth quarter (Q4) encompassed the day before the final exam, covering the period from day 208 to the day before the final exam of the course.

3.3.2. Feature construction

To enhance the accuracy of the prediction process, new features were generated in this study:

1- Total_precourse_activities: This feature combines the values of 20 behavioral features representing student interactions with the VLE before the course starts as shown in Eq. 2. The philosophy behind its calculation is that: Educational Theories: According to Self-Regulated Learning (SRL) Theory, proactive engagement and preparation before the course begins can lead to better academic outcomes as students set goals and plan their learning activities.

$$\text{The total pre course activities} = (\text{BDataPlus} + \text{BDualPane} + \text{BOuCollaborate} + \text{BOuContent} + \text{BOuelluminate} + \text{BExternalQuiz} + \text{BFolder} + \text{BForumng} + \text{BGlossary} + \text{BHomePage} + \text{BOuWiki} + \text{BPage} + \text{BQuestionnaire} + \text{BQuiz} + \text{BBRepeatActivity} + \text{BResource} + \text{BSharedSubPage} + \text{BSubpage} + \text{BUrl}) \quad (2)$$

2- Total_postcourse_activities: This feature combines the values of 20 behavioral features from four different quarters in the original dataset representing student interactions with the VLE after the course starts as shown in Eq. 3. The philosophy behind its calculation is that: Educational Theories: The Engagement Theory posits that ongoing interaction with learning materials and activities is essential for effective learning .

$$\begin{aligned}\text{The total pre course activities}\\ &= (\text{DataPlus} + \text{DualPane} + \text{OuCollaborate} + \text{OuContent} + \text{Ouelluminate}\\ &+ \text{ExternalQuiz} + \text{Folder} + \text{Forumng} + \text{Glossary} + \text{HomePage} + \text{OuWiki}\\ &+ \text{Page} + \text{Questionnaire} + \text{Quiz} + \text{RepeatActivity} + \text{Resource}\\ &+ \text{SharedSubPage} + \text{Subpage} + \text{Url}) \quad (3)\end{aligned}$$

3- Average: It calculate the weighted average of individual assignment scores, where each score is related with a specific weight as shown in Eq. 4. The philosophy behind its calculation is that: Educational Theories: Assessment Theory emphasizes that weighted averages provide a fair representation of student performance by accounting for the varying importance of assignments.

$$\text{Weighted Average} = \frac{\sum(\text{Score} \times \text{Weight})}{\sum \text{Weight}} \qquad (4)$$

4- Engagement: Based on assignment scores and total student activities on a VLE, this feature is calculated. Both academic performance and active participation (clicks) can signify engagement in online learning. The philosophy behind its calculation is that: Educational Theories: The Student Engagement Theory suggests that combining behavioral engagement (activities) and academic engagement (assignment performance) gives a student comprehensive measure of overall engagement as shown in Eq. 5.

$$\text{Engagement} = \frac{(\text{current\_last\_assessment} + \text{Total Student Activities})}{2} \qquad (5)$$

5- Studying: This feature references all the actions about consulting resources. Studying is generated using five features in the original dataset as shown in Eq. 6. The philosophy behind its calculation is that: Educational Theories: According to the Information Processing Theory, frequent and diverse interactions with study materials enhance learning and retention.

$$\text{Studying} = \frac{\text{Resource} + \text{Url} + \text{Page} + \text{Folder} + \text{Dataplus}}{5} \qquad (6)$$

6- Discussing: This feature represents students' communication actions. The philosophy behind its calculation is that: Educational Theories: Social Constructivism emphasizes that learning occurs through social interaction and collaboration as shown in Eq. 7.

$$\text{Discussion} = \frac{\text{ouelluminate} + \text{oucollaborate} + \text{Forumng}}{3} \qquad (7)$$

7- Examining: This feature is about students' evaluation. It is generated based on questionnaire, externalquiz and quiz as shown in Eq. 8. The philosophy behind its calculation is that according to the Formative Assessment Theory mentions that quizzes and questionnaires can play a critical role in the learning process. They provide ongoing feedback to students and teachers about students' understanding and progress. underscores the role of regular evaluations in providing feedback and guiding learning.

$$\text{Examining} = \frac{\text{Questionnaire} + \text{ExternalQuiz} + \text{Quiz}}{3} \qquad (8)$$

8- Working: This feature is about students' navigation through course pages. The philosophy behind its calculation is that based on the Time Management Theory, efficient navigation and use of course resources can be an indication of better time management and organizational skills as shown in Eq.9.

$$\text{Working} = \frac{\text{Homepage} + \text{Subpage} + \text{SharedSubpage}}{3} \tag{9}$$

3.4. Feature selection

To reduce the dimensionality of the feature space, select significant features for the prediction process, and improve prediction accuracy (Moslemi, 2023), a feature selection approach was deployed. Knowing the very significant characteristics of the target class is the task of FS approaches (Yağcı, 2022). The amount of characteristics can be increased or decreased by feature selection, which may have an impact on a model's accuracy. In this work, the Wrapper technique is employed as a conventional feature selection methodology. Heuristic search algorithms and sequential selection are two categories under which wrapper approaches fall (Abdullah and Al-Azawei, 2023). The name sequential selection algorithms come from the way features are added. To find the pertinent characteristics, these techniques employ strategy search. One example of a wrapper approach is sequential feature selection, which starts with an empty set. Features are individually added to a prior subset in each phase. Until the ideal precision is achieved and the right number of features are added, the process is repeated (Abdullah and Al-Azawei, 2023).

3.5. The construction of the ontology model

Semantic Web Rule Language (SWRL) rules were built and generated to forecast students' final performance based on the applied strategy and the developed decision tree (Jabardi and Hadi, 2020). The suggested ontology was constructed using Protégé 5.6. Protégé, developed by the Stanford Biomedical Informatics Research Center, is a free, open-source ontology editor and a framework for knowledge management (Jabardi and Hadi, 2020). In creating a knowledge base, ontology reasoners were utilized to categorize and validate the ontology, examining it to extract information (Bolock, Abdennadher and Herbert, 2021). The classes and attributes of the model were represented in an OWL (Web Ontology Language) ontology, created using the Stanford Protégé ontology editor. Most classes consisted of science and social science subjects (George and Lal, 2019). The OULA class was used as the primary definition for apparent properties, specifying elements such as site, glossary, gender, age band, and imd band.

3.6. Semantic rules

Semantic Web uses the SWRL as a standard guideline language. Using a subset of Rule Markup Language (RuleML) with OWL DL or OWL Lite, SWRL may be used to define rules and logic. Couples of antecedent-consequent rules form under SWRL (Jabardi and Hadi, 2020). The head is the part that comes after the antecedent, which refers to the body (rules). One or more atoms come together to form the head and body. Eq. 10 and 11 are used to determine the syntax of the SWRL rule (Jabardi and Hadi, 2020).

$$B1 \wedge B2 \wedge \ \ldots . \wedge Bn \rightarrow \ H \qquad (10)$$

where, H: head (an atom) and Bn: body (all atoms)

The atoms in SWRL are defined based on Eq. 2.

$$C(j)|O(j,k)|D(j,v)|\ T(v)|Built - Ins\ (f, v1, \ldots, vn) \rightarrow \ Atom \qquad (11)$$

v1... vn = Data kind variable terms or Data kind value names and f = Built-in name. Where T = Data type, C = Class, D = Data type Property, O = Object Property, J, k = Object individual names or object variable names, The realization of semantic web and the encouragement of innovative applications built upon the rules are seen as fundamental pillars of SWRL. Thus, SWRL can be applied to deduce new information from existing facts (Bolock, Abdennadher and Herbert, 2021). Each rule in the domain Ontology is expressed in terms of Ontology concepts and is stored as OWL syntax. Some classification-related decision tree rules are shown in Fig.2.

```
Decision Tree Rules:
 |--- imd_band <= 0.5000000000000000
    |--- AQ1:HomePage <= 0.0878712870180607
       |--- gender <= 0.4583333283662796
          |--- class: 1
       |--- gender >  0.4583333283662796
          |--- class: 1
    |--- AQ1:HomePage >  0.0878712870180607
       |--- AQ1:Glossary <= 0.9583333432674408
          |--- class: 1
       |--- AQ1:Glossary >  0.9583333432674408
          |--- class: 0
```

Fig.2. Some of the decision tree rules

According to significant related features, the important rules with the best results are selected. The rules were converted into SWRL. Then applied Pallet reasoner to infer whether the final performance of a student is pass or fail. The ultimate classification choice is based on the class score. Table 1 lists the categorization requirements for SWRL rules. These decision tree rules can be derived from the decision tree. In a decision tree, each branch can have a set of leaves. The leaves represent the prediction or decision, and the path from the root of the tree to the leaf denotes the process by which the decision was made and the conclusion was reached. Some examples of rules are presented below.

Table 1. The SWRL testing rules are crated from existing facts from science course F-Q1 by using the raw features

Rules	SWRL rule
Rule 1: Path: If imd_band is less than or equal to 0.5 and AQ1:HomePage is less than or equal to 0.0879 and gender is less than or equal to 0.4583, then the class is 1.	CourseA_Students(?s) ^ imd_band(?s, ?imd) ^ swrlb:lessThanOrEqual(?imd, 0.5) ^ AQ1_HomePage(?s, ?hp) ^ swrlb:lessThanOrEqual(?hp, 0.0878712870180607) ^gender(?s, ?g) ^ swrlb:lessThanOrEqual(?g, 0.4583333283662796) -> predicted_final_result(?s, 1)
Rule 2: Path: If imd_band is less than or equal to 0.5 and AQ1:HomePage is less than or equal to 0.0879 and gender is greater than 0.4583, then the class is 1.	CourseA_Students(?s) ^imd_band(?s, ?imd) ^ swrlb:lessThanOrEqual(?imd, 0.5) ^ AQ1_HomePage(?s, ?hp) ^ swrlb:lessThanOrEqual(?hp, 0.0878712870180607) ^ gender(?s, ?g) ^ swrlb:greaterThan(?g, 0.4583333283662796)-> predicted_final_result(?s, 1)
Rule 3: Path: If imd_band is less than or equal to 0.5 and AQ1:HomePage is greater than 0.0879 and AQ1:Glossary is less than or equal to 0.9583, then the class is 1.	CourseA_Students(?s) ^imd_band(?s, ?imd) ^ swrlb:lessThanOrEqual(?imd, 0.5) ^ AQ1_HomePage(?s, ?hp) ^ swrlb:greaterThan(?hp, 0.0878712870180607) ^AQ1_Glossary(?s, ?gloss) ^ swrlb:lessThanOrEqual(?gloss, 0.9583333432674408)-> predicted_final_result(?s, 1)

In Table 1, which presents SWRL rules, the number of SWRL rules is aligned with the number of leaves (classes) in the decision tree. Each branch path that concludes with a class is represented on the left side, while the corresponding SWRL rule for this path is provided on the right side.

3.7. The reasoner

In ontology, Pellet is recognized as the most popular OWL reasoner for implementing SWRL rules and deriving new ontology axioms (Massari et al., 2022). The classification process is one of the most common applications of reasoning, as it enables information to be extracted from ontologies and knowledge bases that is not explicitly provided. A number of essential logical functions, including instance testing, description, and consistency checks, are supported by Pellet Reasoner. In the proposed ontology, Pellet was selected for its ability to define custom built-ins for SWRL and its straightforward functionality when operating with SWRL rules and OWL (Zeebaree et al., 2019).

3.8. Evaluation

To get a general idea of the accuracy of the models, it was critical to measure the efficiency of ontology method. The value and performance of the prediction models were assessed using significant assessment measures, including the f-score, recall, accuracy, and precision. In general terms, these assessment measures may be found in Eq. 12, 13, 14, and 15 (Hussain *et al.*, 2020).

$$\text{Accuracy} = (TP + TN) / TP + TN + \sum FP + \sum FN \quad (12)$$

$$\text{Recall} = \text{TP/TP} + \Sigma\text{FN} \quad (13)$$

$$\text{Precision}=\text{TP/TP} + \Sigma\text{FP} \quad (14)$$

$$\text{F1-measure} = (2 * \text{TP}) / (2 * \text{TP} + \Sigma\text{FN} + \Sigma\text{FP}) \quad (15)$$

Where TN is a true negative, FP is a false positive, TP is a true positive, and FN is a false negative.

4. RESULTS AND DISCUSSION

The purpose of this study is to propose a novel method for enhancing course portability and transferability through the application of ontological engineering and Semantic Web Rule Language (SWRL) rules. To validate the effectiveness of the proposed method, the Open University Learning Analytics (OULA) Dataset was utilized, with a specific focus on science courses. This dataset, detailed in previous studies, comprises 32,593 instances. The data were divided into 70% for training and 30% for testing. Prediction experiments were conducted across two distinct periods: Quarter 1 (Q1), spanning from the start of the course to day 53, and Quarter 4 (Q4), covering days 201 to 207 after the module's official launch. For each science course, two predictions were performed using numerical datasets and decision tree (DT) methods to evaluate the model's performance. To assess the effectiveness of the proposed method, comparisons were made with several state-of-the-art techniques. A summary of the comparative analysis is provided in Table 2.

Our proposed method demonstrates superior accuracy, achieving 97% in predicting student performance, compared to the state-of-the-art techniques outlined in the Table2. This indicates not only the effectiveness of our approach but also its potential for enhanced portability and transferability across different courses.

4.1 Ontological-based model using science courses

The predictability of models is evaluated across courses within the same domain, considering four distinct scientific course modules (C, D, E, and F). The prediction model from course F is tested through using dataset from other courses namely C, D and E which are in the same domain. In the first phase, the prediction accuracy is obtained based on the original features of the dataset as a baseline especially for Q1 and Q4.

As shown in Table 3 below around 7753 out of the 8473 students were successfully classified as pass with an accuracy of 0.92% in the first phase of the classification findings for Q1. Even so, about 720 learners were labeled as fail. Meanwhile, 356 out of 5167 fail students were classified as pass. Table 4 includes the findings for Q4, the findings show that about 7801out of the 8473 learners were correctly classified as pass to achieve 0.93% accuracy, in view of the fact that 672 were classified as fail. At the same time, about 272 out of 5167 fail students were classified as pass. In Q4, ontology classifiers functioned significantly better than Q1. The confusion matrix shows the findings of the first stage see Tables 3 and 4.

Table 2: Comparative Analysis of Prediction Accuracy

Technique	Method	Dataset	Accuracy/F1 Score	Other Metrics	Notes
Daud *et al.*, 2017	Support Vector Machine (SVM)	Demographic Data	F1 Score: 0.867	Location, natural gas usage, etc.	Strong performance
Umer et al., 2018	KNN, NB, RF, LDA	Assignment Scores, VLE Data	70% (RF)	Assignment scores, engagement data	Consistent accuracy but not portable across courses.
Hussain et al., 2018	J48, JRIP, DT, CART, NB	Social Science Courses	88.52% (J48)	Various algorithms	High accuracy in specific courses but limited portability.
Soni *et al.*, 2018	NB, DT, SVM	Recent Performance Data	83.33% (SVM)	20 selected features	Good accuracy but less interpretable
Our Method	Ontological-based Model with SWRL	Science Courses	97%	Enhanced portability, semantic adaptability	Outperforms traditional methods, high accuracy across courses.

Table 3: The confusion matrix of science courses based on the course F-Q1 rules with original features only.

Pass and Fail Classification		Actual values	
Predictive value		Pass	Fail
	Pass	TP=7753	FP=720
	Fail	FN=356	TN=4811

Table 4: The confusion matrix of science courses based on the course F-Q4 rules with original features only.

Pass and Fail Classification		Actual values	
Predictive value		Pass	Fail
	Pass	TP=7801	FP=672
	Fail	FN=272	TN=4895

The second phase prediction the accuracy is obtained base of the original features and feature generation, the integration of a random search optimizer, and the application of a feature selection approach, as demonstrated in Tables 5 and 6.

Nearby 8103 out of the 8473 students were successfully classified as pass with an accuracy of 0.96% for Q1. Nevertheless, about 370 learners were labeled as fail. Meanwhile, 135 out of 5167 fail students were classified as pass. For Q4, the results show that about 8182 out of 8473 learners were correctly classified as pass to achieve 0.97% accuracy, whereas about 291 were classified as fail. Meanwhile, about 94 out of 5167 fail students were classified as pass. In Q4, ontology classifiers performed significantly better than Q1. The confusion matrix shows the findings of the second stage see Tables 5 and 6.

Table 5: The confusion matrix of science courses based on the course F-Q1 rules in the case of wrapper feature selection and the optimizer technique

Pass and Fail Classification		Actual values	
Predictive value		Pass	Fail
	Pass	TP=8103	FP=370
	Fail	FN=135	TN=5032

Table 6: The confusion matrix of science courses based on the course F-Q4 rules in the case of wrapper feature selection and the optimizer technique

Pass and Fail Classification		Actual values	
Predictive value		Pass	Fail
	Pass	TP=8182	FP=291
	Fail	FN=94	TN=5073

Based on the outcomes of Table 6, it is clear that the accuracy for ontology classifiers outperformed previous literature after employing, using the generated attributes, and random search optimizer. The accuracy of ontology increased significantly in comparison to the results in Table 3 and Table 4. In the second phase, the accuracy was enhanced from 0.92% and 0.93% to 0.96% and 0.97% for the science courses.

Previous studies have been primarily focused on machine learning classification methods, such as Decision Trees, Support Vector Machines (SVM), and Random Forests. Strong performance has been demonstrated by these methods in various predictive tasks. However, a novel perspective is introduced in this approach by leveraging ontology engineering combined with Semantic Web Rules, resulting in an accuracy of 97%. This high accuracy indicates that the ontology-based method is not only competitive with traditional machine learning classifiers but also offers additional advantages in model interpretability.

The integration of decision trees with ontologies allows the strengths of both approaches to be combined. High predictive accuracy has been offered by decision trees, which are well-established in machine learning, while ontologies have been shown to enhance the model's adaptability and interpretability. By converting significant decision tree rules into the Semantic Web Rule Language (SWRL), the predictive power of machine learning is leveraged alongside the reasoning capabilities of ontologies. This combination results in a system that is both accurate and understandable.

Ontology reasoners, such as Pellet, were applied to determine student performance outcomes (pass or fail). This integration of reasoning tools with decision trees and ontologies enables a more nuanced and adaptable prediction framework. Through the use of reasoners, the model not only achieves high accuracy in outcome prediction but also provides a rationale for its predictions, which is essential in educational applications where transparency and interpretability are required.

4. CONCLUSIONS

This research was aimed at enhancing the portability and transferability of predictive models for students' performance. New features were created alongside the raw features. An ontological method was developed using SWRL rules, the OWL language, and a reasoner. To ensure accurate responses to requests, the reasoner was employed to execute all OWL ontology queries. For classifying students’ final performance, the Pellet reasoner was utilized. The findings demonstrated that the portability of the models was greatly enhanced through the proposed ontology. This indicates that ontological models from a single source course can be applied to several target courses without sacrificing forecast accuracy. Overall, it was shown that adding new features enhances the performance of the proposed models. Feature selection and random search optimization techniques were also applied, which further improved the findings. Experimental results demonstrated that, over two quarters, the feature selection technique primarily identified the generated features. Consequently, the highest recorded accuracy for ontology predictions was 97%. In the second quarter (Q4) of the course, out of 8182 out of 8473 learners were correctly classified as pass to achieve 0.97% accuracy, whereas about 291 were classified as fail. Meanwhile, about 94 out of 5167 failed students were classified as pass. Despite these promising findings, certain limitations remain. First, the development of additional new features may further enhance the ontology prediction process, potentially improving portability results. Second, the research model was based solely on one dataset, so additional datasets are required to confirm the research findings and improve generalizability

5. REFERENCES

Abd El-Rady, A. (2020) ‘An ontological model to predict dropout students using machine learning techniques’, in 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS). IEEE, pp. 1–5.

Abdullah, S.R.A. and Al-Azawei, A. (2023) 'Enhancing the Early Prediction of Learners Performance in a Virtual Learning Environment', in National Conference on New Trends in Information and Communications Technology Applications. Springer, pp. 252–266.

Aslam, N. et al. (2021) 'An Improved Early Student's Academic Performance Prediction Using Deep Learning', International Journal of Emerging Technologies in Learning (iJET), 16(12), pp. 108–122.

Bolock, A. El, Abdennadher, S. and Herbert, C. (2021) 'An ontology-based framework for psychological monitoring in education during the covid-19 pandemic', Frontiers in Psychology, 12, p. 673586.

Buddhika, M., Karunarathne, B. and Nanayakkara, V. (2024) 'Predictive Modeling for Enhancing MOOC Completion Rates: A Case Study', in European Conference on e-Learning, pp. 43–51.

Chweya, R. et al. (2020) 'A literature review of student performance prediction in E-learning environment', Journal of Science, Engineering, Technology, and Management ISSN, pp. 7858–9989.

Daud, A. et al. (2017) 'Predicting student performance using advanced learning analytics', in Proceedings of the 26th international conference on world wide web companion, pp. 415–421.

Fadhil, G., Abed, I. and Jasim, R. (2021) 'Genetic algorithm utilization to fine tune the parameters of PID controller', Kufa Journal of Engineering, 12(2), pp. 1–12.

George, G. and Lal, A.M. (2019) 'Review of ontology-based recommender systems in e-learning', Computers and Education, 142. Available at: https://doi.org/10.1016/j.compedu.2019.103642.

Gheisari, M. et al. (2023) 'Data Mining Techniques for Web Mining: A Survey', in Artificial Intelligence and Applications, pp. 3–10.

Grivokostopoulou, F., Perikos, I. and Hatzilygeroudis, I. (2014) 'Using semantic web technologies in a web based system for personalized learning AI course', in Proceedings - IEEE 6th International Conference on Technology for Education, T4E 2014. Institute of Electrical and Electronics Engineers Inc., pp. 257–260. Available at: https://doi.org/10.1109/T4E.2014.36.

Hamim, T., Benabbou, F. and Sael, N. (2021) 'An Ontology-based Decision Support System for Multi-objective Prediction Tasks', International Journal of Advanced Computer Science and Applications, 12(12), pp. 183–191. Available at: https://doi.org/10.14569/IJACSA.2021.0121224.

Hamzah, N.A. and Dhannoon, B.N. (2021) ‘The detection of sexual harassment and chat predators using artificial neural network’, Karbala International Journal of Modern Science, 7(4), p. 6.

Hussain, M. et al. (2018) ‘Student engagement predictions in an e-learning system and their impact on student course assessment scores’, Computational intelligence and neuroscience, 2018.

Hussain, M. et al. (2020) ‘Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Score’, 7, p. 100033. Available at: https://doi.org/10.1016/j.eswax.2020.100033.

Iqbal, M. et al. (2024) ‘Personalized and adaptive e-learning systems for semantic Web: a systematic review and roadmap’, International Journal of Web Information Systems [Preprint].

Jabardi, M. and Hadi, A.S. (2020) ‘Twitter fake account detection and classification using ontological engineering and semantic web rule language’, Karbala International Journal of Modern Science, 6(4), p. 8.

Kuzilek, J., Hlosta, M. and Zdrahal, Z. (2017) ‘Open university learning analytics dataset’, Scientific data, 4(1), pp. 1–8.

Lee, C.S., Cheang, P.Y.S. and Moslehpour, M. (2022) ‘Predictive analytics in business analytics: decision tree’, Advances in Decision Sciences, 26(1), pp. 1–29.

López-Zambrano, J., Lara, J.A. and Romero, C. (2022) ‘Improving the portability of predicting students’ performance models by using ontologies’, Journal of Computing in Higher Education, 34(1). Available at: https://doi.org/10.1007/s12528-021-09273-3.

Massari, H. et al. (2022) ‘Ontology-based decision tree model for prediction of cardiovascular disease’, Indian J. Comput. Sci. Eng, 13(3), pp. 851–859.

Merchant, A. et al. (2022) ‘Predicting Students’ Academic Performance in Virtual Learning Environment Using Machine Learning’, ICPC2T 2022 - 2nd International Conference on Power, Control and Computing Technologies, Proceedings [Preprint]. Available at: https://doi.org/10.1109/ICPC2T53885.2022.9777008.

Mohammed, H., Kareem, S. and Mohammed, A. (2022) ‘A COMPARATIVE EVALUATION OF DEEP LEARNING METHODS IN DIGITAL IMAGE CLASSIFICATION’, Kufa Journal of Engineering, 13(4), pp. 53–69.

Moslemi, A. (2023) ‘A tutorial-based survey on feature selection: Recent advancements on feature selection’, Engineering Applications of Artificial Intelligence, 126, p. 107136.

Papadakis, S. (2023) ‘MOOCs 2012-2022: An overview’, Advances in Mobile Learning Educational Research, 3(1), pp. 682–693.

Qasrawi, R. et al. (2021) ‘Predicting School Children Academic Performance Using Machine Learning Techniques’, Advances in Science, Technology and Engineering Systems Journal, 6(5), pp. 8–15. Available at: https://doi.org/10.25046/aj060502.

Qiu, F. et al. (2022) ‘Predicting students’ performance in e-learning using learning process and behaviour data’, Scientific Reports, 12(1), pp. 1–15.

Soni, A. et al. (2018) ‘Predicting student performance using data mining techniques’, International Journal of Pure and Applied Mathematics, 119(12), pp. 221–227.

Sultana, J., Rani, M.U. and Farquad, M.A.H. (2019) ‘Student’s performance prediction using deep learning and data mining methods’, Int. J. Recent Technol. Eng, 8(1S4), pp. 1018–1021.

Tan, P.-N., Steinbach, M. and Kumar, V. (2006) ‘Introduction to data mining, Pearson education’, Inc., New Delhi [Preprint].

Tkachenko, K. et al. (2024) ‘Ontological Approach in Modern Educational Processes’, Cybersecurity Providing in Information and Telecommunication Systems 2024, 3654, pp. 88–97.

Ullah, M.S. et al. (2023) ‘Analyzing students’e-learning usage and post-usage outcomes in higher education’, Computers and education open, 5, p. 100146.

Umer, R. et al. (2018) ‘A learning analytics approach: Using online weekly student engagement data to make predictions on student performance’, in 2018 International conference on computing, electronic and electrical engineering (ICE Cube). IEEE, pp. 1–5.

Yağcı, M. (2022) ‘Educational data mining: prediction of students’ academic performance using machine learning algorithms’, Smart Learning Environments, 9(1), pp. 1–19.

Zeebaree, A. et al. (2019) ‘Designing an ontology of E-learning system for duhok polytechnic university using protégé OWL tool’, J Adv Res Dyn Control Syst Vol, 11, pp. 24–37.

Zhou, X. et al. (2024) ‘Recent advances in data mining and machine learning for enhanced building energy management’, Energy, p. 132636.