

An Ontology-based Decision Support System for Multi-objective Prediction Tasks

Touria Hamim, Faouzia Benabbou, Nawal Sael
Laboratory of Modeling and Information Technology
Hassan II University, Faculty of Sciences Ben M'sik, Casablanca, Morocco

Abstract—Student profile modeling is a topic that continues to attract the interest of both academics and researchers because of its crucial role in the development of predictive or decision support systems. It provides platforms to build intelligent systems such as e-orientation, e-recruitment, recommendation, and prediction systems. The purpose of this research is to propose an ontology-based decision support system that can be used for multi-objective prediction tasks such as prediction of failure/abundance, orientation or decision-making. Two major contributions are proposed here: a new domain ontology that models the profile of a student and a system that is based on this ontology to perform multiple prediction tasks. The proposed approach relies on the efficiency of the ontology to ensure semantic interoperability and the benefits of machine learning techniques to build an intelligent system for a multipurpose decision support objectives. The proposed system uses Decision Tree algorithm (C5.0), but other machine learning models can be added if they prove to be more efficient. Furthermore, the performance of the developed method is computed using performance metrics and achieved 83.6% for accuracy and 81.9% for recall.

Keywords—Profile modeling; student; ontology; machine learning; academic domain

I. INTRODUCTION

In the educational field, student can be described over different information that changes over time and which constantly evolves. The profile model is the way to represent and cover the different dimensions describing accurately different features of the student such as personal, academic, social, psychological information and some others [1]. The reliability or quality of the profile description is very crucial to have efficient understanding of students. Profile modelling has the advantage of encompassing several aspects of the student that can be exploited for different purposes such as course recommendation, orientation, outcome prediction, recruitment, etc.

The education sector is a great field of promise for the uses of artificial intelligence. Artificial intelligence (AI) has the capacity to meet some of the greatest challenges facing the field of education today to develop innovative teaching and learning practices. Machine Learning (ML) can be used in many ways in the field of education, either for adaptive learning which, depending on the abilities and learning mode of each student, allows to choose personalized techniques and optimized techniques to the individual scale, or for improving student performance by identifying the cause of the problem and helping to remedy it and the institutions themselves can

identify their weaknesses and find areas for improvement to maximize their students' results, or for predicting student success or failure. ML techniques have also shown their power to help students choose their path, where based on the data, the system can suggest a student to work in the industry or sector that best suits him, guidance can therefore be based on these results to better guide those who are struggling to find their way. Machine learning-based systems are used to design complex models and algorithms that lend themselves to prediction or decision support. These models allow researchers, scientists, engineers and analysts to “produce reliable and repeatable decisions and results” and to uncover “hidden information” by taking advantage of historical relationships and trends in data. To learn, these systems receive huge amounts of data, which they then use to learn how to perform a specific task. The quality and size of this dataset is important to building a system capable of performing the task assigned to it with precision. Researchers turned their attention to the possibilities of standardizing the representation of knowledge. Ontology is one of the approaches that allow concepts to be represented explicitly, it determines the concepts that exist or may exist in the area of interest [2-3].

In addition to representing knowledge, a machine learning-based system must encode the knowledge into a form that can be processed efficiently. The Ontologies are currently among the most talked models in Knowledge Engineering aiming to establish representations through which machines can manipulate the semantics of information, integrate new concepts according to the evolution of the system and use different data sources. Ontologies offer knowledge sharing facilitated by the use of a common conceptualization (vocabulary and semantics) and the adoption of a standard ontological language. The integration of machine learning (ML) techniques in ontologies makes it possible to enrich and broaden the context of use, from an ontology that offers a common conceptualization, to an ontological system of decision support. Several research studies combine ontologies with machine learning techniques for different purposes [4-5]. This combination can be used for the enrichment of the data available to a Machine Learning model, and this comes from the fact that ontology, offers in addition to raw data, a whole chain of associations and relations between data.

This paper presents two major contributions. The first one is the creation of a new Generic Student Modeling ontology (GSMonto) that aims to describe several student features in different levels while remaining extensible and scalable. The second contribution is the exploitation of GSMonto to build an

ontology-based decision support framework for multi-objective tasks. The objectives are numerous such as to offer adapted learning content, to predict student failure, success or dropout, to propose student orientation, or recommendation, etc.

This article is organized as follows. In Section 2, the authors present the background concept of ontology and machine learning techniques. Section 3 is a comparative study of the different researches dealing with student profile modelling using ontologies and dealing with the integration of Machine Learning with ontologies. The proposed generic student model ontology (GSMonto) is detailed in Section 4. Section 5 develops the proposed ontology-based machine learning system for student profile. Section 6 presents an experiment of the proposed system. At last, the authors give a discussion, conclusion and perspectives.

II. BACKGROUND CONCEPT

A. Ontology

The definition of ontologies is inherited from a philosophical tradition which is concerned with the science of "Being". Today, it means the "science of beings" that is to say the set of objects recognized as existing in a domain. It is a structured set of concepts that make sense of information [6]. Its primary objective is to model a body of knowledge in a given field.

The advantage of ontologies is the separation of knowledge. The ontological knowledge being separated, it can be reused in several applications, and these re-uses (total or partial) can form the basis of interoperability between different systems. For example, integrating an ontology into a ML based system therefore makes it possible to formally declare a certain amount of knowledge used to characterize the information managed by the system and to be based on these characterizations and the formalization of their meaning in order to automate data processing tasks. Ontologies are employed as a form of representation of knowledge in Artificial Intelligence, semantic web, software engineering, biomedical domain, and information architecture. Among the constraints of the use of ontologies is their creation difficulty as well as the visualization limits and the difficulty of finding ready-made ontologies to meet user needs.

There are several knowledge representation languages, as RDF (Resource Description Framework) [7] and RDF Schema [8] which have tried to solve the problem of the absence of the semantics of XML schemas by associating simple semantics with identifiers, RDF and RDFS were designed to be as generic as possible, this simplicity of language is also accompanied by an insufficient expressiveness for the description of complex situations. The OIL (Ontology Inference Layer) [9] and DAML (DARPA Agent Markup Language) languages were developed to fill the gaps of RDF, OIL allows defining classes and relations and a limited number of axioms and DAML intervenes to allow agents to share semantics. These two languages were then merged to give a DAML+OIL language that is based on the RDF and RDF Schema languages by enriching them with new primitives [10]. Based on the DAML + OIL language, OWL was defined based on the basic primitives defined by RDF schemas. However, far from being

a simple extension of RDF, OWL provides all the semantics necessary for the description of knowledge especially for publishing and sharing resources on the semantic web, structuring them in an understandable and standardized way, and making them accessible by adding Meta information. For this, OWL is chosen to lead this study, given that OWL has more powerful means of expressing meaning and semantics than XML, RDF, and RDF-S. In addition, OWL allows information to be gathered from distributed sources, including allowing the import of information from other ontologies. OWL is developed as an extension of the RDF vocabulary and is derived from the DAML + OIL ontology language [11]. OWL has three increasingly expressive sub-languages. OWL Lite is a sub-language that supports users who mainly need a classification hierarchy and simple constraints, which makes the calculation time of inference processes limited. The advantage of OWL Lite is that is both easier to grasp (for users) and easier to implement (for tool builders). The disadvantage is its restricted expressivity. OWL DL, a sub-language of OWL that supports maximum expressiveness needs while guaranteeing the completeness of calculations and decidability necessary for reasoning systems. The advantage of OWL DL is that it permits efficient reasoning support but loose full compatibility. And OWL Full sub-language which gives the user maximum expressiveness, but there is no guarantee as to the completeness and completion of the procedures inference, the advantage of OWL full is its maximum expressiveness without sacrificing computational completeness and the inconvenient is that it is so powerful in expressiveness that it became undecidable. OWL DL is the language chosen to drive the proposed approach.

B. Machine Learning Techniques

Machine learning is an artificial intelligence (AI) field that enables systems to learn and improve automatically from the experience itself without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn on their own. In general, two main types of machine learning algorithms are used today: supervised learning and unsupervised learning. In supervised Learning all data is tagged and algorithms learn to predict the outcome of the input data. Unsupervised Learning uses an unlabeled data set, the machine is then asked to create its own responses. It thus offers answers from analyses and grouping of data.

In the proposed system of this paper, Decision Tree (DT) is used, which is an algorithm that estimates a target concept by a tree representation, where each internal node corresponds to an attribute, and each terminal node (or leaf) corresponds to a class. It is widely known and used in many fields to aid the decision-making. In the academic field, it was shown in author's previous study [12], that it is counted among the algorithms that give the best performance for the prediction of the academic performance of the students. There are several automatic algorithms for building decision trees like ID3, C4.5, C5.0 and CART, etc. ID3 (Iterative Dichotomiser 3) was developed by Ross Quinlan. It can be applied only on the nominal characteristics. It is used for ranking. Therefore, if the data contains continuous characteristics, then discretization should be applied. C4.5 is an extension of ID3 by Ross Quinlan

that can be applied on all types of features. Among the improvements of C4.5 regarding ID3, is the transformation of continuous (digital) characteristics into nominal characteristics dynamically, features without values are ignored when calculating entropy and information gain and pruning trees after creation. C5.0 is a commercial and vastly improved version of the C4.5 algorithm, which applies in large databases. Among the improvements that are added in the C5.0 algorithm is the addition of new data types such as dates and the improvements in efficiency, memory and data processing speed. And CART algorithm, which is similar to that of C4.5 with a few differences like supporting regression, representing the decision tree by a series of binary divisions leading to terminal nodes which can be described by a set of specific rules. The attractive presentation of the CART tree makes it easy to interpret.

III. RELATED WORKS REVIEW

A. Ontology Construction in Educational Domain

In several research studies, ontologies were used to master the information resources of learners, and to facilitate their organization and exploitation. Ontologies is now one of the most important bases of the Semantic Web approaches in the educational field, whether for online or traditional learning. When building a model in adaptive e-learning, several questions arise, such as which information will serve to represent a better the learner? Which formalism to choose for representing and managing the learner model? To answer these questions, authors in [13] proposed an ontology-based approach for the representation of learner profile and learning styles to simply use them for a personalized E-learning and to allow greater flexibility and reusability. The proposed approach collects personal information about the learners, their learning styles, prerequisites, preferences, objectives, online behaviors, etc.

In adaptive e-learning, the student information can be traced and used by systems to provide adaptive content, the authors in [14] addressed the involvement of digital library in the e-learning process and proposed a student model that adopts technologies, applications and standards from the Semantic Web by using OWL ontology language based on six main classes: Personal data, Background, Motivation state, Learning goal and Preferences. In adaptive learning, the content is adjusted to the learner's profile to respect their learning style. In [15], authors proposed to establish an ontological relationship between the formation of learner models with adaptive learning systems. As well, the authors in [16] presented an OWL learner ontology that supports personalization based on three learner style models: Felder-Silverman, Honey-Mumford and Kolb. While an ontology-based approach with six learning style model to adapted learning systems was presented in [17]. In [18], authors proposed a Framework for adaptive learning ontology to retrieve learning resources according to the learner style and knowledge level. The ontology captures information about learner's personal information, prior knowledge, and learning styles. In [19], proposed an ontology to model learners enrolled in distance learning. The proposed ontology arranges learner model characteristics into facets. The Learner class is the key concept of our hierarchy and it includes all specific details

regarding learners. It's associated with the corresponding subclasses through has Profile, has Education and has Personality.

One of the biggest challenges of adaptive e-learning systems is learner modelling. To create a model that meets the requirements, the authors in [20] proposed a novel adult learner's knowledge model using ontologies and rule reasoning. The proposed model takes into account different elements of the learner's knowledge. In [21] a ubiquitous lifelong user model ontology called LifeOnto is proposed, which meets the requirements of adaptive learning systems. Authors in [22] proposed a model based on OWL-DL ontology language that can provide support for recommended activities and personalization of educational context in Adaptive Educational Systems (AES) by grouping the chosen characteristics into four classes: Personal Information, behavior, context and progress/knowledge. An ontology-based learner modelling approach is proposed in [2] to adapt learning contents to learner. Four main classes were proposed, namely: Personal data, context, cognitive data and activity data.

In order to adapt the learning profile to the learning environment, the authors in [23] focus on the following behavioral analysis and evaluation, the detection of learning styles, the development of the learner's profile that takes into account the knowledge, preferences and attitude of the learner for learner profile modeling. Authors in [24] proposed an approach based on the semantic student profile to predict learning preference of the students based on their learning interest and style. With the advent of e-learning, even school orientation starts to be done remotely. The author in [25], proposed a framework of an ontology system called for personalized course recommendation. The approach aims to integrate the information from multiple sources based on the hierarchical ontology similarity with a view to enhancing the efficiency and the user satisfaction and to provide students with appropriate recommendations. The proposed user profile consists of two main parts. The first part is the personal attributes and education attributes of the user and the second is the user's rating of the previously recommended course.

The links between higher education and the world of work are the most controversial. Most of the controversy revolves around the mismatch between higher education opportunities and the needs of the world of work. A better understanding of the relationship between education and the world of work helps to pinpoint the reality of the problems that higher education encounters. Studying the gap between the results of higher education and the needs of industry was the author's objective in [26], by establishing an ontological relationship between the skill requirements of market occupations and the profile of learners of higher education to ensure continuous alignment between student profiles and industry. The classes used by the author to establish the ontology of the student profile are divided into five sub-models: Common model, Education model that represents the education profile, Student model, Application model and Occupation model. Table I presents a second comparison of the research studies dealing with the construction of ontologies in the academic domain for different objectives. Each research study is presented in a row of the table, including the reference, publication year, ontology language used, the tool and the main ontology model classes

used (to define and categorize the different concepts used, to a standardization of concepts is proceeded).

B. Integration of Machine Learning Techniques with Ontologies in Different Domains

ML learning integration with ontologies has proven to be successful in many decision-support systems. For this, artificial neural network methods, logical rules based techniques like decision tree, mathematical functions based ones like SVM, probabilistic methods like naive Bayesian classifier and some others, are used over ontologies concepts or data. The author in [27], used an environmental ontologies of lakes with the K-means clustering algorithm to group lakes according to the average nitrogen concentration into two groups (PoorIn and RichIn). In [28] and [29], the authors proposed an artificially intelligent predictive model for a manufacturing network by developing an ontology model based on decision tree algorithm. In [30], the main goal of the research is enhancing ontology matching by using techniques coming from different fields such as ML, Information Retrieval and Graph Matching to discover correspondences between semantically related entities of ontologies by transforming the ontology matching task into a classification task in ML (match or not match category) using Decision Tree J48 model. Authors in [5], proposed an ontology-based decision tree where the principle was using characteristics of the elements and the relations between them to find the feature super-class with the highest

information gain instead of using a single vector of characteristics in the model. These classes are used as decision on tree node to obtain more information on the preferences of the user. The relation between ontologies and ML techniques can be described as a reciprocal benefit relationship. In addition to the advantage of the application of ML techniques within the ontology for decision support, the ontology also carries a benefit for Machine Learning techniques especially in the data processing due to its organized structure which is especially the case for Text classification issue. As for the authors in [4] who used a Human Disease Ontology, and tried to carry out a classification problem with and without the use of ontology. The authors found that the ontology based classification stands at a higher level than the classification without ontology by using various ML classifiers.

The study of the research studies cited in the educational field shows that researchers are interested in constructing the ontology for a specific objective, either to adapt the content of e-learning to the profile of each student, or to predict the drop-out or the performance of the student, or for recommendation or guidance. According to this objective, the system is modelled to meet the given purpose. In what follows, a multi-objective decision support system is proposed, developed from an ontology which covers the different concepts of the student with the integration of ML techniques.

TABLE I. COMPARISON OF CONSTRUCTED ONTOLOGIES FOR EDUCATIONAL PURPOSES

Ref	Ontology language	Tool	Ontology Main Classes
[13]	-	-	Personal data, Prerequisite, Preference, learning style, online behavior, social data, psychological data.
[14]	OWL	Protégé	Personal data, Learning style, knowledge, Course information
[15]	-	-	Personal data, knowledge, Online behavior, skills, Interaction, Activity.
[16]	OWL	Protégé	Personal Data, Learning style, Education.
[17]	OWL	Protégé	Learning style
[18]	OWL	Protégé	Personal data, Learning style, knowledge, Course information
[19]	OWL	Protégé	Personal data, Knowledge, Learning style, Cognitive data, Preferences, Motivation, Education, Goals.
[20]	-	-	Personal data, Social data, Knowledge, Cognitive data, Personality, Psychological data.
[21]	OWL	Protégé	Aptitude, Bloom taxonomy, cognitive capability, disability, personality, stereotype, degree, language, history, learner, learning approach, learning style, plan.
[22]	OWL-DL	-	Personal data, Behavior, Context, Progress/Knowledge.
[2]	OWL	Protégé	Personal data, context, Cognitive data, Activity data.
[23]	OWL	Protégé	Personal data, Knowledge, Behavior, Interaction, Skills, Activity, Preferences.
[24]	OWL	Protégé	Personal data, Social data, Education, learning style
[25]	OWL	Protégé	Personal data, Education, Skills
[26]	OWL	Protégé	Common model, Education model, Student, Application model, Occupation model.

IV. GSMONTO: GENERIC STUDENT MODEL ONTOLOGY

A. *GSMonto Creation*

Based on state of the art and different data sources, the authors build a domain ontology that covers the different aspects of the student needed in building intelligent systems, in the context of keeping up lifelong learning processes, go beyond guaranteeing interoperability between different educational systems or applications via the web, capable of satisfying users educational needs, and that can be used for different purposes (prediction of failure, dropout, orientation, etc.). OWL DL language is used for its ability to be distributed across many systems and its scaling for Web needs with reasoner support. The generic student model ontology (GSMonto) is based on expert knowledge and documentation on the educational field and mainly covers different classes that can be scalable for future perspectives. The ontology creation goes through two fundamental stages, which are the acquisition and the modelling of knowledge. During the construction of ontology, a semantic reasoner is used to deduce logical consequences from a set of facts or asserted axioms (Hermit and Pallet are two Protégé reasoners examples).

The proposed ontology defines a set of 12 upper level classes, namely:

- Student: the main class of the proposed ontology.
- Personal Identity: defines the student in a unique way (first name, family name, personal address, etc.).
- Social Identity: Derived from the belonging of a student to a social group (Nationality, marital status, etc.).
- Digital Identity: Student Identity in Social Media and Web World (digital signature, twitter account, etc.).
- Family Background: Describes the family size, parents' education, parents' job, family structure (if the student has no parent, single parent or two parents), etc.
- Personality: Encompasses information about the psychology (if the student has a normal psychology or suffers from addiction, depression or anxiety), skills and personality type of the student (intuitive, extraverted or introverted).
- Professional Experiences: Internships and jobs that has been done by the student.
- Physical Limitation: student's physical limitation as visual, verbal, hearing, amputation, paralysis, etc.

- Knowledge Profile: Records if the student has general, theoretical or practical knowledge in a specific topic.
- Learning Profile: Encompasses information about the student's learning in presential and online learning platforms, like learning style, interaction preferences (practice, example or principle oriented), learning media (audio, text or video oriented), etc.
- Academic Background: Detail about the formal education that the student has received including high school background, university background (graduate education, certified formations or PhD formation).
- Cognitive Profile: Describes the range of mental processes relating to the manipulation of the information like memory level, intelligence level, etc.

Fig. 1 describes the schematic representation of the proposed ontology. Each class can include data type properties to connect a single subject with some form of attribute data, and object properties to provide the relationships between two individuals from given class.

B. *GSMonto Scalability and Instanciation*

To ensure the update and the scalability of the ontology, an enrichment of the latter has to be done periodically. This enrichment can be defined following either the knowledge of experts in the educational field, or by the collection of a dataset which carries new concepts from different sources, whether from learning management systems (Moodle, Blackboard, etc.), Enterprise Resource Planning (like Apogee, XML, etc.) or traditional databases. Then, a mapping process is necessary, to convert dataset components or new discovered concepts to the corresponding ontology components. When it comes to the data source, a Metadata analysis is done, to obtain a description of the data. This metadata analysis will be used to create mapping rules, which, for each database component, generate the correspondence in ontology components with the objective of creating an ontological model from a database. If the concepts drawn from the mapping rules and those of the constructed ontology are aligned, an automatic instantiation of the constructed ontology is done in the third. Nevertheless, if the mapping rules shed light on new concepts that don't exist in the ontology already established, and which are supposed to be important, an update must be done so that the ontology is up to date, as shown in Fig. 2 that describes the process of updating and instantiating the ontology.

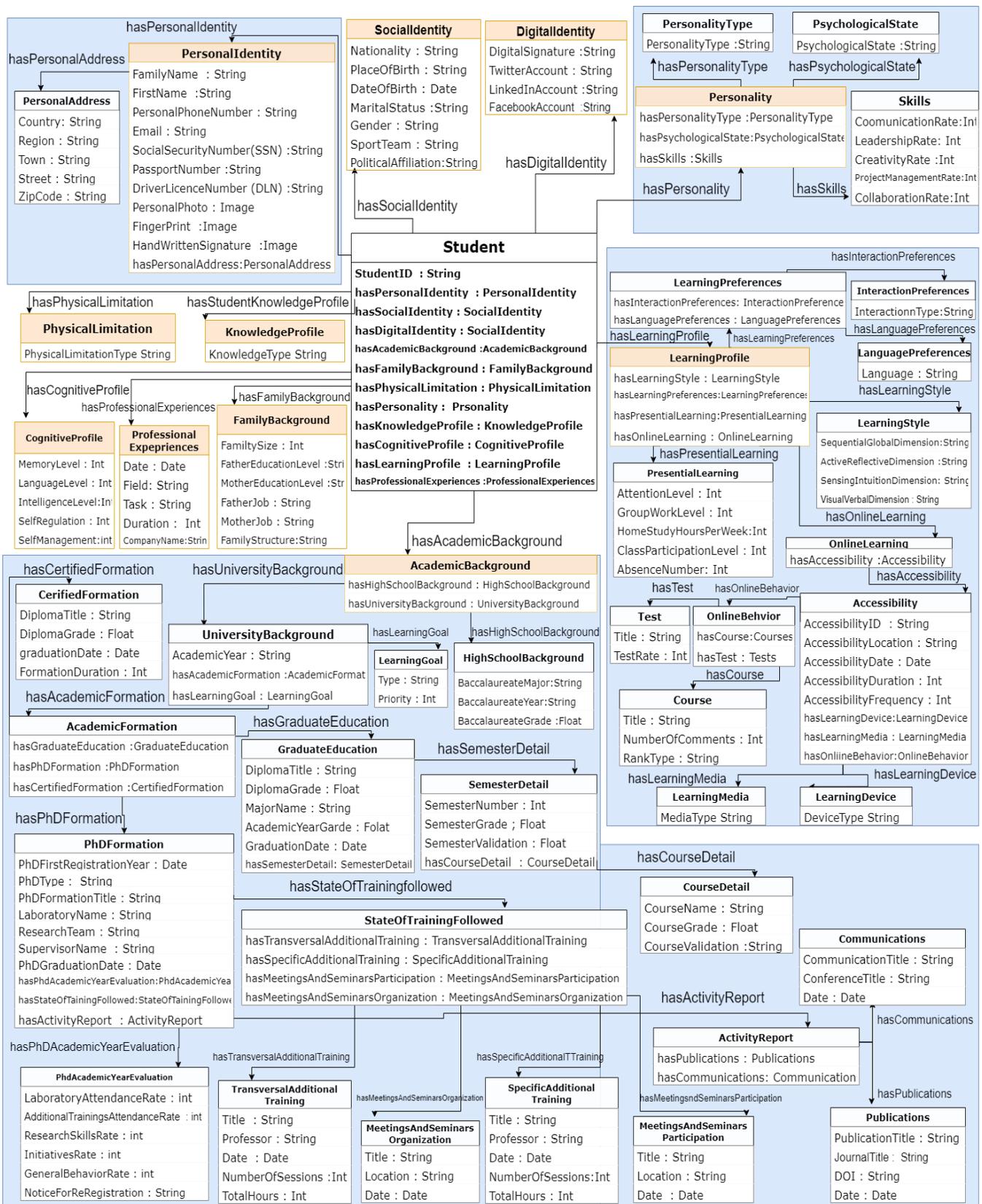


Fig. 1. GSMonto Overview.

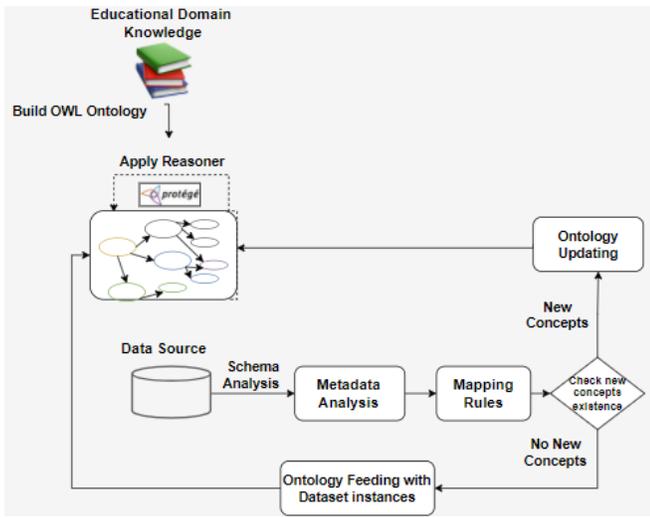


Fig. 2. Ontology Update and Instantiation Process.

V. PROPOSED ONTOLOGY BASED MACHINE LEARNING SYSTEM FOR STUDENT PROFILE MODELING

This paper proposes an ontology based Machine Learning system using an ontology that covers the different aspects of the student. The proposed system process of the can be divided into four levels. The first level concerns the exploitation of the constructed ontology. According to the objective (prediction of

failure, dropout, orientation or other objective) and to the available data, a selection of relevant concepts of the ontology is made to generate a sub-ontology. The second level is the conversion of the created sub-ontology to a dataset using mapping tools. These are based on the use of defined rules to transform the concepts of ontology (classes, individuals, data type properties, etc.) to their equivalents dataset that can be used for the application of ML techniques.

In the third level, the ML process is applied on the generated dataset, starting with data pre-processing, in the case of the presence of anomalies or incorrect values that compromise the quality of the dataset, knowing that the initial ontological presentation of the data already offers the advantage of avoiding the majority of inconsistencies, conflicts and contradictions. Then, a feature selection is carried out to reduce the number of input variables and the computational cost of modelling and, in some cases, to improve the performance of the model. In addition, data partitioning is followed to prepare for the application of the appropriate ML technique. The final level consists in converting the ML algorithm results into Semantics Web Rule Language (SWRL) to be easily integrated into the ontology, thus enriching its expressive power and increasing knowledge about individuals, so the ontology will be more consistent and can include in addition to knowledge, integrated predictive models. Fig. 3 shows the proposed ontology based ML system process for student profile modelling.

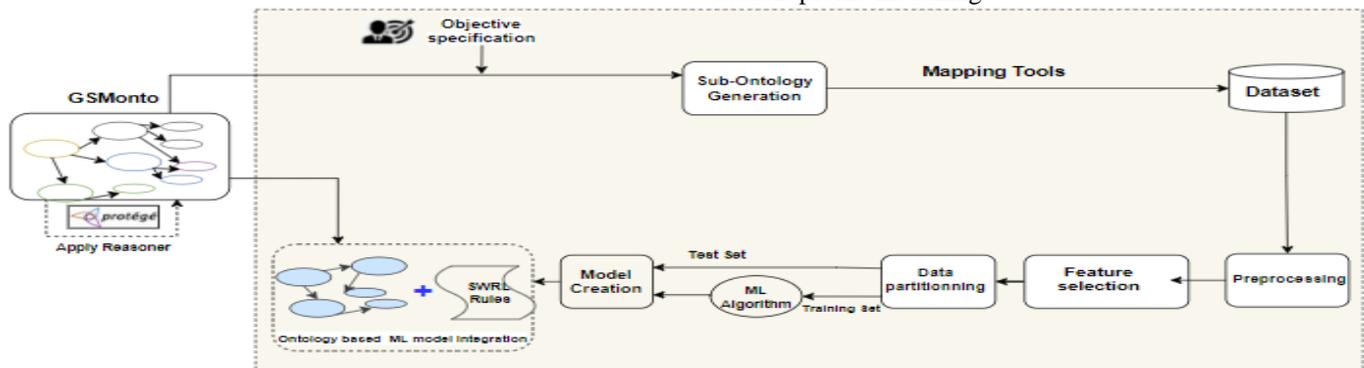


Fig. 3. The Proposed Ontology-Based Decision Support System for Multi-objective Prediction Tasks.

VI. EXPERIMENT AND RESULTS

In this section, a case study that illustrates the application of the proposed system described in the previous section is presented. The aim of this case study is to apply the proposed system in real situations, whether they are new and/or complex or to extend knowledge. The implementation process of the GSMonto ontology is accomplished with the aid of Protégé tool, which is a free and open source ontology editor for building intelligent systems.

A. Data Source and Metadata Analysis

The dataset used is extracted from the Apogee (Application For Organization and Management of Teaching and Students) database and transferred to a spreadsheet, including 20 academic performance features of 3911 student [31].

One feature relates to the validation result of the academic year (100) and two features relate to the validation of each

semester of the two semesters (110 and 120), and the rest of the attributes give the validation results of the courses.

B. Ontology Updating and Feeding

In this case study, no new general context is detected, but detailed sub concepts must be presented in the established ontology, these are the courses studied in each semester. The semester class is divided into two classes (semester 1 and semester 2) and each course column title from the Spreadsheet is transformed to a subclass of each Semester class. To carry out this mapping, Cellfie is used, which is a Protégé Desktop plugin for importing spreadsheet data into OWL ontologies with the intermediary of the mapping rules. The updating process of the ontology includes information about the new courses. For example, a transformation rule expression can be defined to take the name of the spreadsheet cell G1 (column name SemesterOne) and declare an OWL named class that is a subclass of an existing YearResultDetail class which defines

that the SemesterOne class is a subclass of the class YearResultDetail , as follows:

Class: @G1 SubClassOf: YearResultDetail

Fig. 4 illustrates the proposed ontology with the update undergone in relation to new concepts in Protégé.

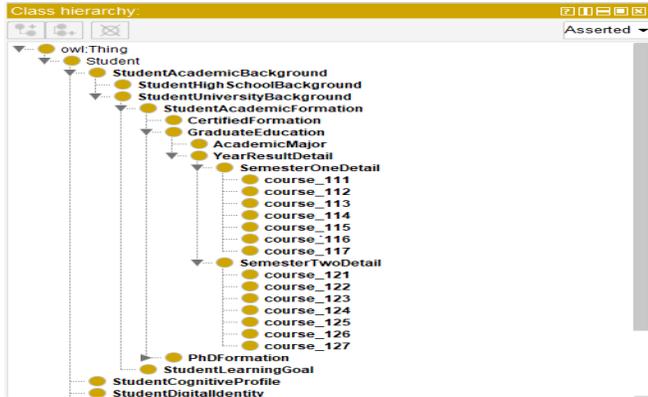


Fig. 4. GSMonto overview after the Updating Process as Displayed in Protégé.

After updating the new concepts under the GSMonto, Pallet reasoned is used to ensure the compatibility between the concepts. Finally, the ontology feeding is carried out.

C. Machine Learning Integration

The Third step of the proposed process concerns the integration of ML within the ontology created.

An example of prediction concerns the same dataset with which the authors instantiated the proposed ontology. The validation of the academic year is the class to be predicted. The independent variables used for the prediction concerns the validation result of the first semester and its related courses. The algorithm used in this case study is C5.0. R software is used for the implementation of the model. The model performances achieved are 83.6% in accuracy and 81.9% in recall. To implement the proposed model within the ontology, the rules deduced from the decision tree are converted into SWRL rules, using a Python code that maps each rule in the decision tree to its equivalence syntax on Protégé SWRL. The generated decision tree rules are converted into SWRL rules. An example the conversion method is presented in the Fig. 5.

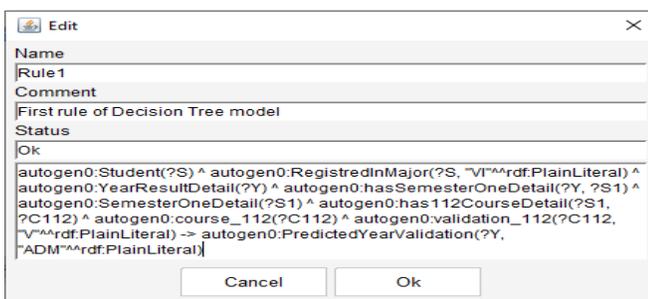


Fig. 5. SWRL Rules Implementation.

For each new student registered with the information of his academic performance in the first semester, the model gives the academic year result prediction. Fig. 6 shows the prediction result for a student: PredictedYearValidation: “ADM”, which means that the student will succeed in the academic year according to the model prediction.

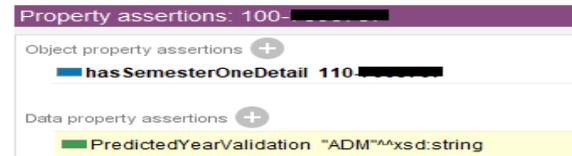


Fig. 6. Prediction Results view in Protégé.

VII. DISCUSSION

In this study, the authors were able to implement a generic ontology that dissects the majority of students’ characteristics. The authors were also able to bring together the benefits of standardization of the concepts provided by the ontology with the benefits of machine learning techniques to meet several predictive tasks. As an experiment, the decision tree algorithm was used for predicting student performance, and the results of the algorithm were transformed into SWRL rules to build an ontology-based decision support system.

VIII. CONCLUSION

The authors have presented a generic ontology proposed for modelling the student profile, which differ from the existing ontologies by its generic aspect that can be adapted to several objectives in the educational field. The authors also proposed a system that combines the proposed ontology with machine learning, using an algorithm based on decision trees and SWRL rules to achieve several objectives such as prediction of failure/abundance, orientation or decision-making. As a future perspective, authors plan to optimize the update process of the ontology by automatically detecting and integrating new concepts. Another goal is to implement other machine learning techniques to meet other objectives and to benefit from the AHP (Analytical Hierarchy Process) technique implemented in a previous paper to propose a module in the system dealing with the missing data as in the case of school dropout.

REFERENCES

- [1] T. Hamim, F. Benabbou, and N. Sael, “Student profile modeling: an overview model”, in Proceedings of the 4th International Conference on Smart City Applications, New York, NY, USA, oct. 2019, p. 1-9.
- [2] L. Akharraz, A. El Mezouary, and Z. Mahani, “LMonto: An Ontology-Based Learner Model for Technology Enhanced Learning Systems”, International Conference on Advanced Information Technology, Services and Systems. Springer, Cham, 2018. p. 137-142.
- [3] H. Yago, J. Clemente, D. Rodriguez, and P. Cordoba, “ON-SMMILE: Ontology Nandwork-based Student Model for Multiple Learning Environments”, Data Knowl. Eng., vol. 115, p. 48-67, mai 2018.
- [4] S. Malik and S. Jain, “Semantic Ontology-Based Approach to Enhance Text Classification”, CEUR Workshop (Vol. 2786, pp. 85-98).
- [5] A. Bouza, G. Reif, A. Bernstein, and H. Gall, “SemTree: ontology-based decision tree algorithm for recommender systems”, 2008.
- [6] N. Guarino and P. Giarandta, “Ontologies and Knowledge Base”, Towards very large knowledge bases, 1995, p. 1-2.

- [7] O. Lassila, R. R. Swick, W. Wide, and W. Consortium, "Resource Description Framework (RDF) Model and Syntax Specification". 1998.
- [8] V. Christophides, "Resource Description Framework (RDF) Schema (RDFS)", in Encyclopedia of Database Systems, L. LIU and M. T. ÖZSU, Éd. Boston, MA: Springer US, 2009, p. 2425-2428.
- [9] D. Fensel, I. Horrocks, F. van Harmelen, D. McGuinness, and P. F. Patel-Schneider, "OIL: Ontology Infrastructure to Enable the Semantic Web", IEEE intelligent systems, 16(2), 38-45.
- [10] D. L. McGuinness, R. Fikes, J. Hendler, and L. A. Stein, "DAML+OIL: an ontology language for the Semantic Web", IEEE Intell. Syst., vol. 17, no 5, p. 72-80, sept. 2002.
- [11] McGuinness, D. L., and V.Harmelen, F "OWL Web Ontology Language Overview", W3C recommendation, 2004, vol. 10, no 10, p. 2004.
- [12] T. Hamim, F. Benabbou, and N. Sael, " Survey of Machine Learning Techniques for Student Profile Modeling", International Journal of Emerging Technologies in Learning, vol. 16, no 04, p. 136, févr. 2021.
- [13] S. Bourekkache, O. Kazar, M. Abik, S. Tigane, and L. Kahloul. "Ontology based approach for representing the learner profile and learning styles". In : 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS). IEEE, 2019. p. 1-6.
- [14] D. Paneva, "Use of Ontology-Based Student Model in Semantic-Oriented Access to the Knowledge in digital libraries", In proc. of HUBUSKA Fourth Open Workshop (pp. 31-41).
- [15] S. Ulfa, D.B . Lasfeto and C. Kurniawan., "Modelling The Learner Model Based Ontology In Adaptive Learning Environment". Journal of Disruptive Learning Innovation (JODLI) (2019).
- [16] B. Ciloglulig and M. M. Inceoglu, "A Learner Ontology Based on Learning Style Models for Adaptive E-Learning", in Computational Science and Its Applications – ICCSA 2018, mai 2018, p. 199-212.
- [17] A. E. Labib, J. H. Canós, and M. C. Penadés, "On the way to learning style models integration: a Learner's Characteristics Ontology", Comput. Hum. Behav., vol. 73, p. 433-445, août 2017.
- [18] A.MUNASSAR and A.ALI, "Semantic web technology and ontology for E-learning environment". Egyptian Computer Science Journal, 2019, vol. 43, no 2, p. 88-100.
- [19] O.Zine, A.Derouich and A.Talbi, "IMS Compliant Ontological Learner Model for Adaptive E-Learning Environments". International Journal of Emerging Technologies in Learning, 2019, vol. 14, no 16.
- [20] A.Abyaa, M.IDRISSI, and S.Bennani, "An adult learner's knowledge model based on ontologies and rule reasoning". In : Proceedings of the Mediterranean Symposium on Smart City Application. 2017. p. 1-6.
- [21] D. Nurjanah, "LifeOn, a ubiquitous lifelong learner model ontology supporting adaptive learning", in 2018 IEEE Global Engineering Education Conference (EDUCON), avr. 2018, p. 866-871.
- [22] H. N. M. Ferreira, T. Brant-Ribeiro, R. D. Araujo, F. A. Dorca, and R. G. Cattelan, "An Automatic and Dynamic Student Modeling Approach for Adaptive and Intelligent Educational Systems Using Ontologies and Bayesian Networks", in IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2016.
- [23] A.Korch, N. El Amrani El Idrissi and L. Oughdir. "Modeling and Implementing Ontology for Managing Learners' Profiles". International Journal Of Advanced Computer Science And Applications, 2017, vol. 8, no 8, p. 144-152.
- [24] T. Sheeba and R. Krishnan, "Semantic Predictive Model of Student Dynamic Profile Using Fuzzy Concept", Procedia computer science, 2018, vol. 132, p. 1592-1601.
- [25] M.Ibrahim, Y.Yang, D. L.Ndzi., G.Yang and M. Al-Maliki. "Ontology-based personalized course recommendation framework". IEEE Access, 2018, vol. 7, p. 5180-5199.
- [26] H. Gasmi and A. Bouras, "Ontology-Based Education/Industry Collaboration System", IEEE Access, vol. 6, p. 1362-1371, 2018.
- [27] M. Stocker, M. Rönkkö, F. Villa, and M. Kolehmainen, "The Relevance of Measurement Data in Environmental Ontology Learning", in International Symposium on Environmental Software Systems. Springer, Berlin, Heidelberg, 2011. p. 445-453.
- [28] X.-B. Tang, G.-C. Liu, J. Yang, and W. Wei, "Knowledge-based Financial Statement Fraud Detection System: Based on an Ontology and a Decision Tree", Ko Knowledge Organization, 2018, vol. 45, no 3, p. 205-219.
- [29] Z. M. A. Khan, S. Saeidlou, and M. Saadat, "Ontology-based decision tree model for prediction in a manufacturing network", Production & Manufacturing Research, 2019, vol. 7, no 1, p. 335-34.
- [30] D. H. Ngo, "Enhancing Ontology Matching by Using Machine Learning, Graph Matching and Information Retrieval Techniques", (Doctoral dissertation, Université Montpellier II-Sciences et Techniques du Languedoc).
- [31] N. Sael, T. Hamim, and F. Benabbou, "Multilevel Hybrid System Based on Machine Learning and AHP for Student Failure Prediction", International Journal of Computer Science and Network Security, Vol. 19 No. 9 pp. 103-112, 2019.