BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **24**, No 2
Sofia • 2024 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2024-0016

A Recommender System for Educational Planning

Mustapha Skittou, Mohamed Merrouchi, Taoufiq Gadi

Laboratory Mathematics, Computer and Engineering Sciences (MISI), Hassan First University of Settat, Faculty of Sciences and Technics, Morocco
E-mails: mustapha.skittou@gmail.com merrouchi37@gmail.com gtaoufiq@yahoo.fr

Abstract: *Knowledge-based recommender systems have always had their privileged place among all Decision Support Systems (DSS), given their advantage on several points over other techniques. Our paper presents a framework implementing a hybrid form of Rule-Based Reasoning and Case-Based Reasoning (RBR-CBR), to address the rarely discussed domain of educational planning. The system has been tested and presented outstanding results with a high accuracy, which will benefit educational planners' decision support. We have also developed a dedicated application for this project to visualize the results obtained.*

Keywords: *Recommender system, Knowledge-based system, Rule-based reasoning, Case-based reasoning, Hybrid system, Educational planning.*

1. Introduction

Recommender Systems (RS) have undergone considerable development in the present century. This is clearly due to the huge upswing in information and communication technologies, with the advent of artificial intelligence and the Big Data era.

A recommendation system is an exceptional decision-making support tool [1]. It can be used to recommend solutions to decision-makers to compare incidental situations, by studying the specific features of items, users, or the history of previous experiences [2]. There are several methodologies used to build a recommendation system. Collaborative Filtering (CF) [3] is one of the most widely used, especially in e-Commerce [4] and e-Learning [5], and is based on the paradigm that if two users share an interest in previous products, they are likely to like the same products in the future [6]. According to Breese and other authors (1998), CF can be broken down into two main trends: Memory-based and Model-based. A second type is the Content-Based recommender System (CBS), which profiles the user by considering the various metadata describing him or her, to suggest the items most suited to his or her specific needs [7].

Finally, the Knowledge-Based Recommendation System (KBRS) is a privileged decision-support tool that gathers information from different sources and is very close to the knowledge domain [8]. KBRS avoids several very worrying problems in CF or CBS, notably the Cold Start or Sparsity problem. There are two main types of knowledge-based reasoning: Rule-Based Reasoning (RBR) and Case-Based Reasoning (CBR). RBR responds to rules established by an expert in the knowledge domain [9], and looks for associations between elements in the knowledge base [10]. Given its simplicity and speed, the RBR represents the ancestor of knowledge-based reasoning systems. However, since the development of intelligent machine learning techniques, it has begun to give way to other expert systems. CBR, for its part, is a technique widely used in decision support systems, and is very rigorous, since it is based on concrete data and exploits the results of experiments in such situations [11]. Some consider it to be an improvement on Content-based RS [12] to address various types of uncommon situations by benefiting from our learning history.

Through this present research, our main contribution can be summarized as follows:

- Develop a hybrid system framework model [13] integrating case-based reasoning and rule-based reasoning. We adopted a sequential approach, first using the RBR to fill in the column of the primary data set containing the proposed conclusions. This is followed by the CBS module, which takes over the key-driven base to produce the most suitable recommendations in response to new cases.
- Demonstrate the effective application of our approach by using machine learning techniques with various classification algorithms to recommend educational planning actions in response to situational educational cases. These techniques are applied to an original database we have assembled from the Moroccan educational information system.
- The system we proposed gave good results and performed very well for the problem we addressed. Accuracy reached 100% for the training set and the test set. This suggests that the system will have the potential to stipulate very suitable decisions for decision-makers in the field of educational planning.

Our paper is organized as follows. Section 2 presents a literature review of the various works that have generally adopted our approach's tendency. Section 3, describes the materials used in our approach to build our RS and the different methods we based on to elaborate our proposed solution. Section 4 then presents the results of our solution implemented to address the problem of educational planning. And finally, in Section 5, the conclusion summarizes our ideas.

2. Related work

Further research has already demonstrated the remarkable usefulness of hybrid approaches for knowledge-based systems, each with its own philosophy relating to the knowledge domain and techniques deployed.

Norzaidah et al. [14] carried out a comparable study between the rule-based and case-based implementation of an intelligent tutoring system using cognitive models. The researchers used the criteria of knowledge representation, learning, search strategy efficiency, user feedback, input incompleteness, and knowledge base expansion. The choice of adopting RBR or CBR is made based on user testing since both techniques have advantages as well as limitations. RBR is challenging for the definition of appropriate rules, and the system becomes slower as these rules multiply. CBR, on the other hand, offers greater flexibility in incorporating user interaction, although a major difficulty is the construction of the case base. Rimuljo Werdiningsih et al. [15] also carried out an experimental study on the classification of childhood illnesses, comparing the use of RBR and CBR. The results showed a wide gap between the accuracy of CBR and RBR. In particular, they used a weighted distance of CBR with the KNN algorithm, and for RBR they opted for Forward chaining.

Ramchand et al. [16] have combined RBR and CBR to define the cloud typology to be adopted. The cloud is either public, private, or hybrid. The aim is to support strategic decisions for business sponsors. CBR is used to highlight additional requirements that describe the potential of a cloud migration action, based on an analysis of previous experience. Then comes RBR to recommend the most suitable for the new case, based on the learning achieved by CBR. This method represents an inverted approach to our idea, but it raises the challenge as well of the availability of a historical base's cases that researchers can use for requirements generation.

The pairing of RBR and CBR techniques has a strong presence in the medical field. Admass and Munaye [17] have developed a hybrid knowledge-based system integrating RBR and CBR for the diagnosis and treatment of Mango disease. First, rules are developed based on expert knowledge and through the application of data mining techniques to raw data. Then, the cases are written using COLIBRI software. The role of the two reasoning techniques is complementary since when a new query is entered, the RBR takes charge of it; if it coincides with the rules library, it responds, otherwise, it passes it on to the CBR module. Renata et al. [18] used the same hybrid approach to propose a model of a decision support system for the medical diagnosis of gastrointestinal cancer. The approach adopted was to make CBR the general method of the system, and to involve CBR in the data retrieval stage to refine the solution.

Autonomous driving is a theme taken up by the research of Maoyuan et al. [19] by sequentially combining RBR and CBR to ensure safe, accident-free driving of vehicles. The RBR will maintain the elements of speed limit, route change, and traffic light sensitivity. For its part, the CBR receives new events, takes into account environmental factors, and produces decisions based on the knowledge base. In addition, the behaviors of the autonomous vehicle are modeled through a highly detailed ontology.

It can be seen that the majority of methods employed combining RBR and CBR techniques can be classified into four modes, either CBR first and RBR last, or RBR first and CBR last. There's also the parallel mode, where RBR and CBR are used simultaneously, and the result is an aggregation between the two outputs of the two techniques. Finally, in some cases, we find that RBR is incorporated within a CBR lifecycle step, leaving the main method of the system to be CBR [18].

3. Materials and methods

3.1. Dataset formulation

The importance of the knowledge base presents a real challenge in its formulation and construction. One of these challenges is access to information, which is often scattered and difficult to collect. In our case, we exploited several operational systems of the Moroccan education system, each of which manages a separate component: MASSAR (School Management System), GRESA (School Portfolio Management System), ESISE (Statistics and Census Information System), and CarteSco (Education Planning System). Most of the data we have found useful for our approach is an aggregation of the databases from the systems presented before. Table 1 shows the most important attributes of our project.

Table 1. Most relevant features

Features	Designation	Type	Values domain
CD_Com	Commune code	Alphanumeric	Code
Id_Student	Student ID	Numeric	Code
CD_Etab	School code	Alphanumeric	Code
Id_Mil	Milieu ID	Numeric	Discrete {1, 2}
Type_Etab	School type	Numeric	Discrete {1, 2}
Id_Genre	Gender ID	Numeric	Discrete {1, 2}
Lib_Formation	Formation level	Numeric	Discrete {1, 2, 3, 4, 5, 6}
Delay_Sco	School delay	Numeric	Continuous
Id_Provenance	Provenance ID	Numeric	Discrete {0, 1, 2, 3, 4, 5, 6}
Id_Handicap	Disability ID	Numeric	Discrete {0, 1, 2, 3, 4, 5, 6}
Final_Note	Average overall Note	Numeric	Continuous
Id_Result	Final result ID	Numeric	Discrete {1, 2, 3}
Appui_Social	Social aid benefit	Numeric	Discrete {1, 2}

The attributes shown in Table 1 represent those we have deemed most relevant to our approach, and as far as possible.

3.2. Preprocessing

Collecting information from several sources often results in inconsistencies in attribute names and data types, as well as difficulties with missing data. To tackle the problem of data fusion and database unification, we opted for the School code “Code Gresa” as the joining attribute, as illustrated in Fig. 1.

The choice of the school’s Gresa code isn’t arbitrary, it’s mainly motivated by the fact that this attribute is ubiquitous in all operational information system databases, and generally has the same name: Cd_Etab. It’s a unique alphanumeric attribute that identifies each school in its own right, and allows us to identify the various information related to the school: Pedagogical structure (students, classes, teachers, subjects, grades...), physical infrastructure (classrooms, administration, grounds, surface area, networking...), and services (social support, boarding school, school transport...). We carried out several pre-processing manipulations, mainly the elimination of several tuples concerning: Free students, Students studying in the original education system, Private school pupils, and Schools with a drop-out rate > 70. All these rows could have an immense effect and bias the model.

As previously stated, we are limited to working with primary cycle data, for reasons of data availability and the limited scope of our analysis. For this reason, we have retained for our data set only schools with nature codes: 200, 201, 202, 203, and 209.

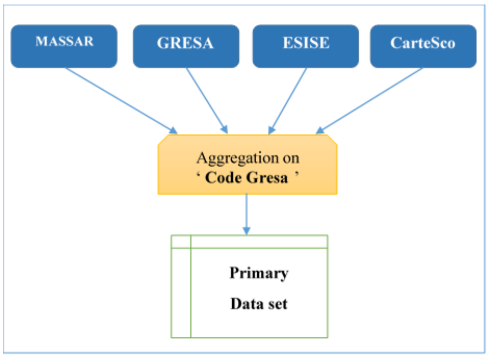

Fig. 1. Educational data sources aggregation

3.3. Indicator-based dataset

To be consistent with the logic of educational planning, and to prepare the database for decision-making purposes, we grouped all the attribute data from the primary data set based on the Gresa school code. This gave our dataset the form shown in Table 2.

Table 2. Features the second step

Feature	Designation	Type
CD_Com	Commune code	Alphanumeric
CD_Etab	School code	Alphanumeric
Id_Mil	Milieu Id	Numeric
Type_Etab	School type	Numeric
Genre_M	Student “Male” number	Numeric
Genre_M	Student “Female” number	Numeric
Total	Total of students	Numeric
Delay_M	Sum of School daly for students “Males”	Numeric
Delay_F	Sum of School daly for students “Females”	Numeric
Total_Delay	The sum of the total School daily	Numeric
Prsco_No	Number of non-schooled students	Numeric
Presco_Yes	Number of pre-schooled students	Numeric
Handicap_Nb	Number of disabled students	Numeric
Admitted	Number of Admitted students	Numeric
Not_Admitted	Number of not Admitted students	Numeric
Dropped	Number of dropped-out students	Numeric
Classes_Cluttered	The sum of cluttered classes	Numeric
Total_Classes	Total number of classes	Numeric
Appui_Social	Social aid benefit	Numeric

Except for the CD_Com and CD_Etab attributes, which are codes for the commune and school, all attributes are in numeric format.

Most of the attributes of the processing dataset we’ll be using in our modeling will be in the form of key indicators of the state of education. Algorithm 1 is used to create these school-based indicators:

Algorithm 1. Calculation of database indicator attributes

Data : Intermediate database
Input : *C* is the school code; *M* is the number of male pupils enrolled in the primary cycle; *F* is the number of female pupils enrolled in the primary cycle; *X_M* is the number of male pupils in the primary cycle age group; *X_F* is the number of female pupils in the primary cycle age group; *D_M* is the sum of the educational delay for a male pupil; *F_M* is the sum of the educational delay of a female pupil; *Presco* is the number of pupils benefiting from pre-school education; *S* is the number of pupils admitted at the end of the year; *R* is the number of pupils who repeated the year; *DR* is the number of dropouts; *Cl* is the number of classes; *E* is the number of overcrowded classes.
Output: Processing base
Parity index for raw enrolment ratio (RER) *t*; Average male educational delay *d_M*; Average female educational delay *d_F*; Preschool enrolment rate *p*; Success rate *s*; Repetition rate *r*; Dropout rate *dr*; Overcrowding rate *e*.

For each tuple in Intermediate database do :
parity_index = calcul_Parity(M, F, X_M, X_F)
average_delay_m, average_delay_f = av_Delay(M, F, D_M, D_F)
average_total_delay = average_delay_m + average_delay_f
presco_rate = rate_Presco(M, F, Presco)
success_rate = rate_Success(M, F, S)
repetition_rate = rate_repetition(M, F, R)
dropout_rate = rate_Dropped(M, F, DR)
clutter_rate = rate_Clutter(Cl, E)
Procedure calcul_Parity(M, F, X_M, X_F) :
TBSF <- X_M/M
TBSM <- X_F/F
t <- TBSF/TBSM
return t

Procedure av_Delay(M, F, D_M, D_F) :
d_M <- (D_M/M) x 100
d_F <- (D_F/F) x 100
return d_M, d_F
Procedure rate_Presco(M, F, Presco) :
p <- (Presco/M+F) x 100
return p
Procedure rate_Success(M, F, S) :
s <- (S/M+F) x 100
return s
Procedure rate_repetition(M, F, R) :
r <- (R/M+F) x 100
return r
Procedure rate_Dropped(M, F, DR) :
dr <- (DR/M+F) x 100
return dr
Procedure rate_Clutter(Cl, E) :
e <- (E/Cl) x 100
return e

The addition of the attributes elaborated by Algorithm 1 will form the quasi-final dataset of our processing. Each indicator reflects one dimension of the phenomena affecting the school.

The final phase in the formulation of our dataset is the feeding of a final "Recommendation" attribute, which will contain recommended conclusions based on the analysis of all the indicator values in the project database. This process is carried out using the following Algorithm 2.

Algorithm 2. Aggregation of recommendations

Data: Intermediate database
Input: *Code* is the school code, *Mil* is the school environment, *Type* is the school type, *Total* is the sum of students in the school, *T* Parity index for the gross schooling rate; *D_M* Average schooling delay for male students; *D_F* Average schooling delay for female students; *P* Preschool schooling rate; *S* Success rate; *R* Repetition rate; *DR* Dropout rate; *E* Overcrowding rate.
Output: List of recommendations *L*.

L : List()
For each tuple in Intermediate database do :
If Mil = 1 do :
If DR > 2.1 do :
If T < 1 do : Extend(Recommendation) = "Gender approach"

```
    If D_M > 22.8 OR D_F > 18.6 do : Extend(Recommendation) = "Pedagogic formation"
    If P < 65.4 do : Extend(Recommendation) = "Preschooling"
    If R > 8.9 do : Extend(Recommendation) = "Tutoring"
    If E > 30 do : Extend(Recommendation) = "School construction. Teachers recruitment"
    If S < 89 do : Extend(Recommendation) = "School program revision"

Else If Mil = 2 do :
If DR > 2.1 do :
    If T < 1 do : Extend(Recommendation) = "Gender approach"
    If D_M > 27.1 OR D_F > 23.4 do : Extend(Recommendation) = "Pedagogic formation"
    If P < 47.4 do : Extend(Recommendation) = "Preschooling"
    If R > 8.9 do : Extend(Recommendation) = "Tutoring"
    If E > 15 do : Extend(Recommendation) = "School construction. Teachers recruitment"
    If S < 89 do : Extend(Recommendation) = "School program revision"
L = add(Recommendation)
```

For our present research, we have taken the 2019-2020 season as the reference period for our analysis. As such, we took as our baseline the averages of educational indicators relating to this period published in the compendium of “Education Indicators” drawn up by the Directorate of Strategy, Statistics and Planning under the Department of National Education.

The Moroccan Ministry of Education’s report on education for the year shows that the overall dropout rate rose slightly to 5%, and 2.1 at the primary level, which is the target cycle for our analysis. Of course, this rate varies according to the dropout phenomenon by environment, gender, level, single age, and type of education. Moreover, the analysis of the indicator values we have calculated for each school is highly variable, and there are a not inconsiderable number of schools with flagrant anomalies. This is particularly true of almost all public-sector schools. And this, from another angle, gives meaning and reason to our approach and the choice of type of efficiency and performance indicators we have adopted.

3.4. Knowledge-based recommender

In many cases, a knowledge-based recommender system is well suited to the needs of educational decision-makers. Normally, to implement a KBRS we need to establish the following two preliminary elements.

- **Domain knowledge.** The field we are investigating is that of planning. This is not an easy field to identify, due to its specific features and characteristics on the one hand, and the need to control its inferences on all aspects of education on the other. The profession of educational planning, through all the processes it undertakes, clearly outlines the real aspects of educational planning. The study applied to these processes reveals a set of data classified by order of importance that we can exploit in our knowledge base. In our case, this has enabled us to establish the key indicators governing these aspects and build a vision of how to implement a decision-support system as an indispensable tool to support decision-makers in today's education system.
- **Knowledge base.** Building a knowledge base is likewise a very tedious task. It requires considerable expertise in the knowledge domain. This exercise is intrinsic

to gathering the most relevant information for our analysis. We have already explained the various stages in the creation of this base, which required several updates, transformations, and manipulations to arrive at a usable base.

The knowledge-based recommender will give more advantage and flexibility to our approach, given the nature of the data we process, as well as the specificities of educational planning processes.

3.5. Ontology modeling

Ontology-based modeling is one of the most powerful ways of representing knowledge. It enables us to enumerate all the components of the knowledge domain under study, in addition to the polytomous links between these elements.

The ontology model adopted is used to structure our data dictionary representing the knowledge domain. The entities of the ontology provide a certain primary classification of the analysis strands of the notion of educational planning: School Performance, School Life, Care Structure, and School Environment. Each of these components has its own set of attributes, forming our Knowledge base. Through these attributes, other indicator attributes developed via Algorithm 1 are declined, which will transform our ontology model into another form in Fig. 2.

Fig. 2. The educational planning ontology

This schema traces new connections with the ontology's sub-entities and objects to further clarify the components of our base. The indicator attributes we have developed will be seen as the path of our development and thinking, as well as the philosophy with which we view the notions of educational planning. So we now have our conceptual map on which to build our recommendation tool, which represents the core of our project and its essence.

3.6. RB-CB hybrid recommendation

Recommendation strategies are many and varied, depending on the starting point of the recommendation process or its purpose, as well as the nature of the method implemented. Each strategy requires the definition of the relationship between Item and User, as some focus on Item parameters and others on User preferences. KBRS adopts a special, orthodox, and efficient logic. It focuses on the Knowledge of the analysis domain, which allows us to take full advantage of human expertise about the problem being addressed, as well as the capabilities of intelligent technologies that enable us to draw up complex and rigorous codes that process considerable volumes of data.

The two main types of KBRS are Rule-based RS and Case-based RS. Each differs from its counterpart in terms of logic, but also about the essential prerequisites for recommendation. And each has its advantages, normally, and its limitations. But for our project, we've opted for a hybrid approach that combines these two great domains and brings them together, to take advantage of the best they have to offer, and since they are stronger this way. In architectural terms, this means that each will be involved at a specific stage of our proposed solution.

- Rule-based reasoning

RBR represents the ancestor of cognitive reasoning techniques. Its philosophy lies in the definition of several rules that are applied to a knowledge base, leading to precise conclusions. Each rule is a coded condition of the form "If.. Then", and its consequence is a decision or conclusion. In our case, we have adopted RBR to complete our knowledge base. Through Algorithm 2, the passage of the key attributes we have chosen through the test of pre-established rules has enabled us to set up a cumulative recommendation logic. In the field of educational planning, each indicator reflects a problematic dimension of the educational system and automatically leads to a specific action or intervention as well. The rule-based reasoning cycle is illustrated in Fig. 3.

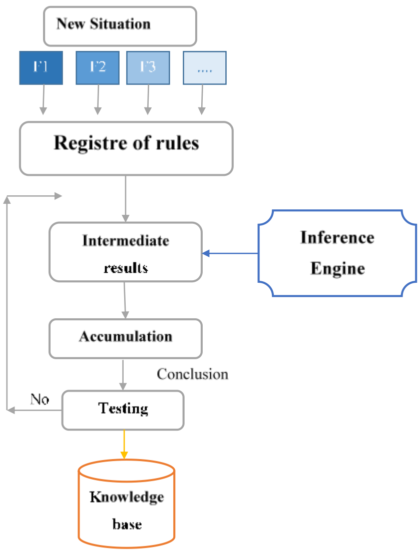

Fig. 3. Rule-based flow chart

The cycle begins with the occurrence of a new situation based on a data model in the knowledge base. Afterward, the register of rules is recalled by applying them to the attributes characteristic of the situation. The subsequent implementation of the inference engine selects the type of chaining adaptable to the situation and produces intermediate results representing elementary responses to the values of the key attributes. These results are then accumulated. We test the general conclusion postulated at the end, and if it is favorable, we incorporate it into the decision/solution part relating to the line in the knowledge base.

- Case-based reasoning

CBR indoctrinates a kind of Content-based recommender system, with a certain specificity [12]. It is based on the use of a Case base, which contains an archive of the actor's experience in a specific domain. The Case base stores all the cases that have been recorded, together with the relevant recommendations that have been made about these cases. Each case represents a specification of the parameters' values of the case attributes, as well as the conclusion stipulated as the answer to the case.

We'll be using the case-based reasoning approach to add more precise performance to our recommendation system. It will also ensure the sustainability and automation of the system since it will rely on an iterative and accumulative base of Case/Solution expertise. The Fig. 4 shows the CBR generative cycle of our project.

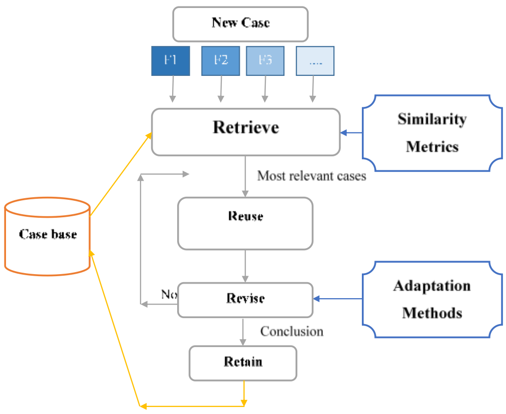

Fig. 4. Case-based flow chart

This part of the CBR represents the heart of our recommendation system. Once the knowledge base has been built, the case-based reasoning cycle enables us to react to new situations as they arise.

3.7. Digitization of conclusions

For each case, the target attribute contains the list of conclusions formulated in response to a specific situation. This represents a major constraint on the data analysis process. The solution lies in recoding these conclusions into a finite numerical code according to,

(1) $$C_k = \{x_1, x_2, x_3, \dots, x_k\},$$

where $\boldsymbol{x}_k$ is the set of conclusions for each case. The transformation of the text values representing the conclusions will consist of replacing them with the code of the intervention group (Table 3).

Table 3. The intervention list

Conclusion	Code
Gender approach	1
Pedagogic formation	2
Pre schooling	3
Tutoring	4
School construction, Teacher recruitment	5
School program revision	6

The processing of text attributes is often a complication for most analysis methods. To make this operation even easier, we opt for recording in digital format. The final form of the target attribute values will be in the form of a global code where the codes of the conclusions stipulated by analysis of the case in question are concatenated, for example: 1234 or 23456.

3.8. Validation

Uncertainty is an ever-present aspect in the processing of any decision-support tool. This is because the stakes are high, and each decision is individually important. Recommendation systems also face this problem, and KBRS adopts several review and test methods to validate their candidate results.

Uncertainty and imprecision can occur at several points in the rule or case-based reasoning cycle. Whether at the level of knowledge domain or rule precision for RBR, as well as the inference techniques used. Or in the choice of key attributes and the application of similarity methods, depending on context and data type.

- Similarity assessment

Calculating similarity between a case that has occurred and historical cases necessarily involves measuring the individual differences between attributes. This logic is well suited to our situation, given that the solution (Recommendation) proposed for each case is a concatenation of the conclusions relating to each key attribute separately. Similarity is seen as the complement of the distance between two entities,

(2) $$(A, B) = 1 - d(A, B).$$

Euclidean Distance

The most commonly used distance is the Euclidean distance. For each attribute list, the distance formula can be presented as follows:

(3) $$\text{EuclideanDistance}(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}.$$

where x_i and y_i are the i-th features of data points x and y, respectively. Euclidean distance is the straight-line distance between two points in an Euclidean space.

Cosine similarity

Cosine is also used to measure the proximity between two data vectors, especially when the magnitude is not very significant,

(4) $$\text{Cosine}(x, y) = \frac{x.y^{\mathrm{T}}}{\|x\|.\|y\|}.$$

- Assessment metrics

To assess the performance of a recommendation system, we use several metrics. These can be classified according to the evaluation domain in terms of accuracy, relevance, and user satisfaction. They generally provide a quantitative measure of the recommendation system's ability to deliver on its objectives.

Accuracy

This is one of the most famous and widely used metrics for judging the veracity of prediction models. It calculates a ratio between the number of correct predictions and the total number of predictions,

$$\text{Accuracy} = \frac{\text{Number\_of\_correct\_predictions}}{\text{Total\_of\_predictions}}. \tag{5}$$

Root Mean Squared Error (RMSE)

Measure calculated on the mean square root of the residual standard deviation by accumulating the square root of the error. It measures the model's degree of accuracy through its sensitivity to the difference between the actual value of the target attribute and its prediction,

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}. \tag{6}$$

Mean Absolute Error (MAE)

Gives the weighted and absolute average of the difference between the predicted and actual value of each element in the database,

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y} |. \tag{7}$$

R-squared (R^2) or Coefficient of determination

It is a statistical measure that indicates the proportion of the variance in the dependent variable that is explained by the independent variables in a regression model,

$$R^2 = 1 - \frac{\sum \text{of\_squared\_regression(SSR)}}{\text{total} \sum \text{of\_squares(SST)}}, \tag{8}$$

$$1 - \frac{\sum\left(y_{i-\hat{y}_i}\right)^2}{\sum\left(y_{i-\bar{y}}\right)^2}.$$

Sensitivity

The Sensitivity, also called "Recall", is the ratio of the number of positive cases that were correct to all the cases that were identified as positive. Mathematically, it looks like this,

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}. \tag{9}$$

Specificity

It is the measure of the proportion of actual negatives correctly predicted by the model,

$$\text{Specificity} = \frac{\text{TN}}{\text{FP+TN}}. \tag{10}$$

3.9. Visualization

Visualization is a tool for presenting the information circuit and encouraging visibility of the idea behind the approach adopted. It is also an interface between the developer and the end-user of the solution.

We opted for an MVC (Model-View-Controller) application, using the Django 2.4 framework based on the Python language, with the Bootstrap MDB package. The application was coded using Visual Studio Code 1.81.1, hosted by the Anaconda 1.10 distribution.

4. Results and discussion

The type of problem our project addresses leads us to choose a knowledge-based recommendation system as the most suitable for our situation. Throughout our research cycle, we used several technologies and techniques to build a solid foundation for the reuse of our approach.

4.1. The proposed Hybrid architecture

In general, knowledge-based recommender systems are favored over other recommender methods in many situations, especially where there is a cold start problem or where there is a need to take into account the business expertise of people in a specific domain.

Our approach consists of a hybrid architecture including the two main types of knowledge-based systems: rule-based reasoning and case-based reasoning (Fig. 5).

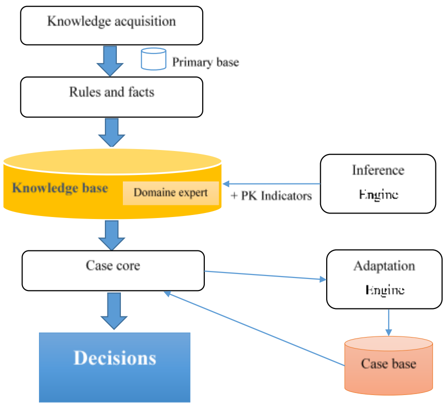

Fig. 5. The proposed framework for hybrid RB-CB reasoning

We start with a raw database collected as explained in Section 3, considering it as the primary database of the project. This is supported by an inference engine programmed according to Algorithm 1 to enrich the base with key performance indicators used in the field of educational planning. In this way, we stabilize an acceptable and quasi-exhaustive knowledge base for use in modeling our solution. This base will form a foundation for the case-based reasoning process. As each new case arises, we look for the most similar cases, use them, and adapt the most optimal solution by interacting with domain expertise. And finally, if it meets the acceptance criteria already established, it is added to the base of cases. The ultimate aim of this recommendation system architecture is to provide educational planners at all levels (central, regional, and local) with a choice of appropriate decisions for the situation at hand.

4.2. Knowledge modeling

The knowledge base consists of 15 descriptive attributes and a 16th that represents the model's target attribute and the case conclusion. To further facilitate data analysis, we encoded the conclusions in a numerical format and then divided them into recommendation groups as illustrated in Table 4.

Table 4. Recommendation groups list

Recommendations group	Number	Code group
12346	100	4
123456	83	3
1234	61	1
12345	60	2
2346	47	8
234	29	5
23456	12	7
2345	10	6
Total of combinations	402	

The number of records used in our project is 402, of which 321 are used for model training and the rest for testing.

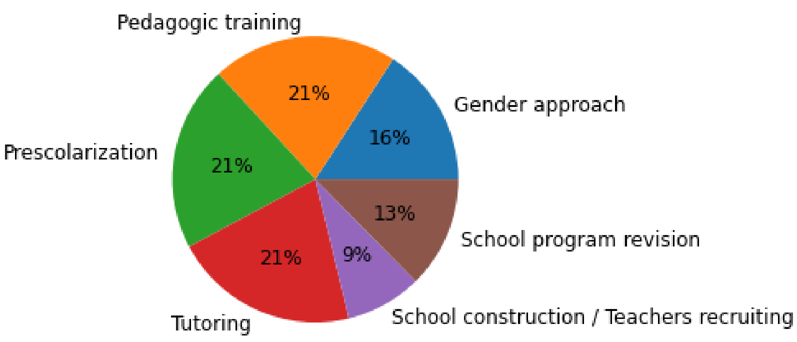

Fig. 6. Conclusions repartition ratio

According to our knowledge base, each case has as its response a group of conclusions made up of at least 3 recommendations, given that the base is multidimensional and groups together various indicators from several aspects of educational planning. Fig. 6 gives us an idea of the representativeness of each conclusion in the set of values in the target column of recommendations.

4.3. Assessment calculation

Depending on the nature of our problem and the choice of methodology, we tried out several techniques to find the most beneficial for achieving our objectives. In particular, we applied a set of machine learning and deep learning algorithms to find the most profitable. The algorithms tested on our knowledge base are Support Vector Machine (SVM), Stochastic Gradient Decent (SGD), K-Nearest Neighbor (KNN), Random Forest (RF), and Neural Network of 3 layers.

To evaluate the performance of each algorithm, we calculated the MAE, RMSE, and R2 metrics as shown in Table 5.

Table 5. Performance measurements for the techniques used

Algorithm	Training set			Test set		
	MAE	RMSE	R2	MAE	RMSE	R2
SVM	1.4205	2.0537	0.3894	1.3456	1.9468	0.3456
SGD	1.6479	2.3620	0.3457	1.9135	2.6874	0.33
KNN	0.9190	1.9519	0.7102	1.66	2.4820	0.4320
RF	0.0	0.0	1.0	0.0	0.0	1.0
Neural network	Loss 0.7394		Accuracy 0.7165	Loss 0.5907		Accuracy 0.7531

The choice of these algorithms is based on our understanding of the problem as one of classification. However, a close look at the results shows that the algorithms that performed best were the most flexible in handling diverse data types through the calculation of distances between the attributes of the new case and the cases in the knowledge base. The algorithm that best predicted the conclusions was Random Forest with an accuracy rate reaching 100%, followed by KNN with 71% accuracy for the training set, and less for the test set.

4.4. ROC curve

To further analyze the accuracy of our chosen RF classifier, we examine the values of the data set description attributes. The prediction accuracy of the target attribute “Conclusion” continues to attain 100% for both the Train set and the Test set.

We investigated the Sensitivity and Specificity variables for the Test set for each class of Label attribute values to determine the prediction deviation recorded, as well as to identify the most problematic classes. Table 6 shows the specificity and sensitivity values for each class.

Table 6. Sensitivity vs specificity measurements

	1	2	3	4	5	6	7	8	Average
Sensitivity	1	1	1	1	1	0	0	1	0.75
Specificity	1	0.9	1	1	0.9	1	1	0.9	0.99

The results indicate that the model registers a very high degree of specificity compared with less sensitivity. This indicates on the one hand that it performs more accurately in correctly predicting negatives. On the other hand, the prediction of positive class cases is far from the expected. Note that for classes 6 and 7, the sensitivity is zero, which is necessarily due to their very low representativeness in the test set.

The ROC Curve in Fig. 7 gives an additional view of the Aires Under the Curves (AUC) for each class of the target attribute. We switched to a One Versus Rest Classifier model to facilitate multi-class processing of True Positive Rate (TPR) and False Positive Rate (FPR). All class curves behave well, the AUC for the majority of classes exceeds 0.99, which is considerable.

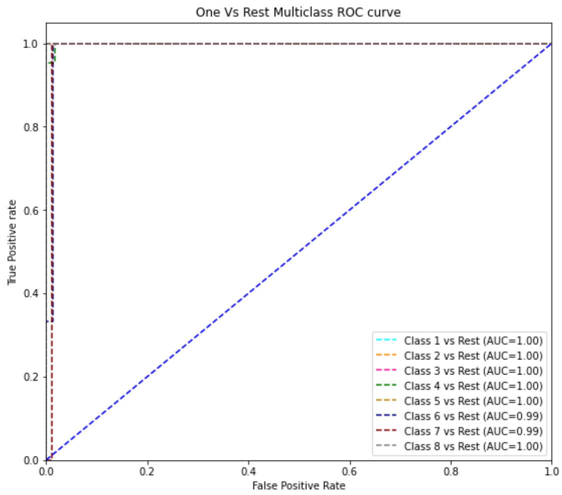

Fig. 7. The one vs rest ROC curve

4.5. Visualization

We have divided the visualization operation into two parts. The first consists of displaying the case base, which is made up of all the cases fed by the algorithmic procedure developed for our approach, as illustrated in Fig. 8.

The second facet of our system is the ability to handle new cases, each of which sets values for each attribute of the form, and a “Recommendation” button when activated allows the most accurate conclusion to be proposed about the criteria already set (Fig. 9).

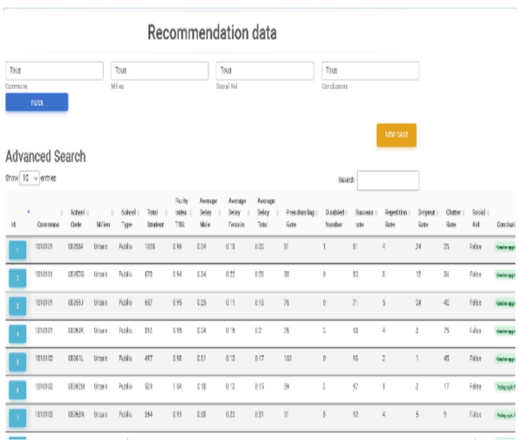

Fig. 8. Display case base

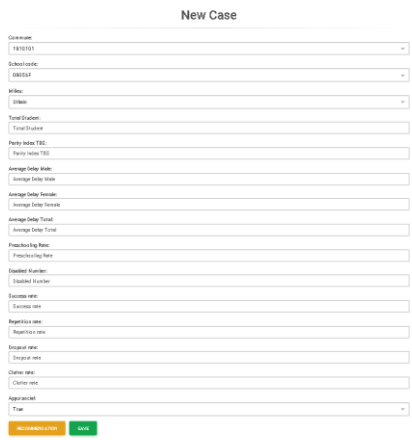

Fig. 9. New case form

4.6. Discussion

To provide educational planners with the support they need, we had to decide on the methodology to adopt. On the one hand, the means available is a very essential element, we’re talking about the dataset as well as the technologies used. On the other hand, the nature of the field to be covered by our contribution – educational planning is a multi-disciplinary and intersecting field, which makes it difficult to pinpoint the issues involved.

The question for the fields of AI and machine learning is always the degree of accuracy of these systems' outputs. Comparisons between different systems objectively show that each type is consistent with different situations. Manel Slokom [20] led a comparison study between recommendation systems applied to original and synthetic data. This demonstrated a change in the behavior of recommendation algorithms depending on the type of data used.

Lahoud et al. [21] are looking for the most effective way of recommending to students the academic paths best suited to their profiles and aspirations. They experimented with a range of standalone systems, and other hybrid combinations. The findings showed that knowledge-based hybridization approaches were the most profitable and yielded the best results.

Given the specificity of education data, the hybrid knowledge-based method is the one that best delivered what was expected of it (Table 7).

Table 7. Recommender systems comparison

Recommender system type	Regulator	Precision
Content-based RS	Cosine similarity	40%
Collaborative Filtering RS	Cosine similarity	20%
Case-based RS	Similarity score	~ 100%

But, to experiment with Content-based or Collaborative filtering RS, we had to perform several grouping manipulations and transform the shape of the dataset to make it processable. For CB RS, we reduced the data frame to a triplet of code_com, code_etab, and conclusions. With the CF RS, we performed considerable dataset manipulation to end up with the model: user_id, item_id, and ratings. And despite all this, our hybrid RB-CB RS knowledge-based approach remained superior on all levels.

5. Conclusion

KBRS continues to demonstrate its value, and its ability to address areas that are not easy for other types of recommendation systems. This close link with the knowledge domain represents their strength and enables us to consolidate the results obtained. In our project, we have unified the best of its two main components, RBR and CBR. Each comes into play at a specific point in our framework, with RBR helping to refine and build the knowledge base, which will later be seen as a case base on which to predict conclusions about future cases. This hybrid architecture has performed very well and demonstrated its usefulness for our educational planning domain, as a decision support tool.

References

1. J a i n, S., A. G r o v e r, P. S. T h a k u r, S. K. C h o u d h a r y. Trends, Problems and Solutions of Recommender System. – In: Proc. of International Conference on Computing, Communication & Automation, Greater Noida, India, 2015, pp. 955-958. DOI: 10.1109/CCAA.2015.7148534.
2. S i n g h, N., S. T r i p a t h i, D. P. S i n g h, P. B h a s k e r. Recommendation Systems in the Big Data Era. – International Journal of Innovative Technology and Exploring Engineering, Vol. **8**, 2019, No 12S3, pp. 80-85.
3. A m u t h a, S., R. V i k r a m S u r y a. An Improved Product Recommender System Using Collaborative Filtering and a Comparative Study of ML Algorithms. – Cybernetics and Information Technologies, Vol. **23**, 2023, No 4, pp. 51-62.
4. M o s i m a, A. M., S. O. O j o, S. A. O d u n a i k e, H. D. M a s e t h e. Framework of Recommendation Systems for Educational Data Mining (EDM) Methods: CBR-RS with KNN Implementation. – Transactions on Engineering Technologies, Singapore, Springer, 2021, pp. 87-98.
5. T r a n, T. D., N. T h a n h-H a i, N. T h a i-N g h e. Novel Approaches for Searching and Recommending Learning Resources. – Cybernetics and Information Technologies, Vol. **23**, 2023, No 2, pp. 151-169.
6. B a n s a l, S., N. B a l i y a n. A Study of Recent Recommender System Techniques. – International Journal of Knowledge and Systems Science (IJKSS), Vol. **10**, 2019, No 2, pp. 13-41. DOI: 10.4018/IJKSS.2019040102.
7. A l h i j a w i, B., Y. K i l a n i. The Recommender System: A Survey. – Int. J. Advanced Intelligence Paradigms, Vol. **15**, 2020, No 3, pp. 229-251.
8. W o r k n e h, A., D. T e f e r i, A. K u m i l a c h e w. Knowledge-Based Decision Support System for Detecting and Diagnosis of Acute Abdomen Using Hybrid Approach. – In: F. Mekuria, E. Nigussie, T. Tegegne, Eds. Information and Communication Technology for Development for Africa. ICT4DA 2019. Communications in Computer and Information Science. Vol. **1026**. Cham, Springer, 2019. DOI: 10.1007/978-3-030-26630-1_6.
9. J u n, G. Network Data Mining Based on Artificial Intelligence Inference Engine. – Microprocessors and Microsystems, Vol. **82**, 2021, 103794. ISSN 0141-9331. DOI: 10.1016/j.micpro.2020.103794.
10. A s i f, M., J. A h m e d. A Novel Case Base Reasoning and Frequent Pattern Based Decision Support System for Mitigating Software Risk Factors. – In: IEEE Access, Vol. **8**, 2020, pp. 102278-102291. DOI: 10.1109/ACCESS.2020.2999036.
11. K a s i e, F. M., G. B r i g h t. Integrating Fuzzy Case-Based Reasoning, Parametric and Feature-Based Cost Estimation Methods for Machining Process. – Journal of Modelling in Management, Vol. **16**, 2021, No 3, pp. 825-847. DOI: 10.1108/JM2-05-2020-0123.
12. S m y t h, B. Case-Based Recommendation. – In: P. Brusilovsky, A. Kobsa, W. Nejdl, Eds. The Adaptive Web. Lecture Notes in Computer Science. Vol. **4321**. Berlin, Heidelberg, Springer, 2007. DOI: 10.1007/978-3-540-72079-9_11.
13. B u r k e, R. Hybrid Recommender Systems: Survey and Experiments. – User Model User-Adap Inter, Vol. **12**, 2002, pp. 331-370. DOI: 10.1023/A:1021240730564.
14. N o r z a i d a h, M. N., A. A h m a d, S. A. H a l i m, A. M. A l i. Intelligent Tutoring System Using Rule-Based and Case-Based: A Comparison, Procedia. – Social and Behavioral Sciences, Vol. **67**, 2012, pp. 454-463. ISSN 1877-0428. DOI: 10.1016/j.sbspro.2012.11.350.
15. W e r d i n i n g s i h, I., R. H e n d r a d i, P. P u r b a n d i n i, B. N u q o b a, E. A n n a. The Efficient Distance Weighted Case Base Rule (DW-CBR) for Early Childhood Diseases Diagnosis. – International Journal of Computing, Vol. **20**, 2021, No 2, pp. 262-269. DOI: 10.47839/ijc.20.2.2174.
16. R a m c h a n d, K., et al. 2018. Towards a Cloud Architectural Decision Framework Using Case-Based Reasoning and Rule-Based Reasoning. – In: Proc. of Australasian Conference on Information Systems, Sydney, 2018. UTS ePRESS. DOI: https://doi.org/10.5130/acis2018.ac.
17. A d m a s s, W. S., Y. Y. M u n a y e. Integrating Case-Based and Rule-Based Reasoning for Diagnosis and Treatment of Mango Disease Using Data Mining Techniques. – Int. J. Inf. Tecnol., 2023. DOI: 10.1007/s41870-023-01587-y.

18. R e n a t a, S., M. P e r k u s i c h, L. S i l v a, H. A l m e i d a, C. S i e b r a, A. P e r k u s i c h. Early Diagnosis of Gastrointestinal Cancer by Using Case-Based and Rule-Based Reasoning. – In: Expert Systems with Applications, 2016. DOI: 10.1016/j.eswa.2016.05.026.
19. M a o y u a n, C., Y. G a o, H. Z h a n, J. H e, W. W e i. An Inference Model of Driving Environment Relationship Based on Rules and Cases. – In: Proc. of SPIE 12079, Second IYSF Academic Symposium on Artificial Intelligence and Computer Engineering, (120791O), 1 December 2021. DOI: 10.1117/12.2622726.
20. M a n e l, S. Comparing Recommender Systems Using Synthetic Data. – In: Proc. of 12th ACM Conference on Recommender Systems (RecSys'18). Association for Computing Machinery, New York, NY, USA, 2018, pp. 548-552. DOI: 10.1145/3240323.3240325.
21. L a h o u d, C., S. M o u s s a, C. O b e i d et al. A Comparative Analysis of Different Recommender Systems for University Major and Career Domain Guidance. – Educ. Inf. Technol., Vol. **28**, 2023, pp. 8733-8759. DOI: 10.1007/s10639-022-11541-3.

Received: 17.01.2024; Second Version: 29.04.2024; Accepted: 28.05.2024