One Aggregated Approach in Multidisciplinary Based Modeling to Predict Further Students’ Education

Milan Ranđelović 1, Aleksandar Aleksić 2, Radovan Radovanović 3, Vladica Stojanović 4, Milan Čabarkapa 5 and Dragan Ranđelović 2,*

1 Science Technology Park, 18000 Niš, Serbia; milan.randjelovic@ntp.rs
2 Faculty of Diplomacy and Security, University Union-Nikola Tesla Belgrade, 11000 Beograd, Serbia; aleksandar.aleksi@fdb.edu.rs
3 Department of Forensic Engineering, University of Criminal Investigation and Police Studies, 11000 Beograd, Serbia; radovan.radovanovic@kpa.edu.rs
4 Department of Information Technology, University of Criminal Investigation and Police Studies, 11000 Beograd, Serbia; vladica.stojanovic@kpa.edu.rs
5 Faculty of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia; cabmilan@etf.rs
* Correspondence: dragan.randjelovic@fdb.edu.rs

Abstract: In this paper, one multidisciplinary-applicable aggregated model has been proposed and verified. This model uses traditional techniques, on the one hand, and algorithms of machine learning as modern techniques, on the other hand, throughout the determination process of the relevance of model attributes for solving any problems of multicriteria decision. The main goal of this model is to take advantage of both approaches and lead to better results than when the techniques are used alone. In addition, the proposed model uses feature selection methodology to reduce the number of attributes, thus increasing the accuracy of the model. We have used the traditional method of regression analysis combined with the well-known mathematical method Analytic Hierarchy Process (AHP). This approach has been combined with the application of the ReliefF classificatory modern ranking method of machine learning. Last but not least, the decision tree classifier J48 has been used for aggregation purposes. Information on grades of the first-year graduate students at the Criminalistics and Police University, Belgrade, after they chose and finished one of the three possible study modules, was used for the evaluation of the proposed model. To the best knowledge of the authors, this work is the first work when considering mining closed frequent trees in case of the streaming of time-varying data.

Keywords: theory of mathematical modeling; applied mathematics; classification and discrimination; multicriteria decision making; linear regression; prediction theory

MSC: 90-10

1. Introduction

In order to clearly present the content of the research covered in this paper, the introduction is divided into two subchapters: ‘Background’, where the problem and the importance and main goal of the solution are presented and ‘Related studies’, where the current state of the research in the field related to the subject of this paper as well as the key publications are given.

The rest of this paper is organized as follows. Section 1 presents the ‘Introduction’, where a description of the considered problem, the background, objectives and existing research gap, as well as our contribution, including the organization of the paper, are given. It gives the authors’ review of the state-of-the-art literature that deals with the topic of interest together with the motivation of the authors to work on this paper. After
that, we present Section 2, ‘Materials and Methods’, as a part of the paper that introduces the problem that the authors solve in this paper, together with the used data material and the methodology that is proposed and adopted in this solution. In Section 3, ‘Results and Findings’, the obtained results are presented using the proposed new model in a concrete case study performed in the Criminal and Police Academy in Belgrade, Serbia. In Section 4, ‘Discussion and Suggestions’, the authors discuss the possibilities of the proposed approach. The end of this paper deals with the concluding remarks in Section 5, which is the ‘Conclusion’.

1.1. Background

Quantification of the importance of relevance of each factor (attribute, criteria) which is included in a usually complex, multidimensional model which describes any problem in many fields of human life, is the subject of research in many fields of sciences. For instance, the scientific disciplines of prediction of the importance of these factors are classification in modern data mining and machine learning, multicriteria decision making and regression as well as correlation analysis in classic mathematics and statistics. It is especially important to pay attention to the need to reduce the dimensionality of these problems to achieve the highest possible accuracy of the model that solves them.

Supervised machine learning, which is used in our proposed model, can be observed as the process where the system improves its performance on the assigned task without additional programming [1]. Supervised learning of understandable knowledge or learning of the concepts (hypothesis, models, classifiers and rules) deals with the induction of logical lawfulness understandable to humans. Since it is also necessary to evaluate the worthiness of implemented knowledge, usually, the set of examples used for supervised learning is divided into two parts: the learning set used for the learning process and the test set used for the testing of acquired knowledge. The basic measurement of the successfulness of acquired (understandable) knowledge is predictive accuracy. It represents the percentage of classification success of new and unconverted examples by using learning rules. Development of some applications of supervised machine learning, above all in construction of the systems based on knowledge and revealing knowledge in information data mining brought to emphasize request for understanding the scientific knowledge. Data mining is the wide-ranging application of technology items relying on several other things such as database technology, statistical analysis as well as artificial intelligence. It gives great commercial value to the penetration of other occupations such as the retail, insurance, telecommunication and power industries use [2].

However, decision making is the process of selection between alternative operational directions to achieve goals and tasks. A multi-criteria decision that we will use in our proposed model which we will propose is one of the most famous branches in decision making with wide application in solving real problems. The presence of various criteria, of which some should be maximized and some minimized, means that the decisions are brought under conflict conditions. These instruments that have to be applied here are more flexible than strict mathematical techniques related to pure optimization.

Analytic hierarchy process-AHP is a mathematical method [3,4] that, compared to the other methods and decision-making techniques, offers the possibility for decision-makers to compare the individual importance of an alternative compared to the others within decision making on an important criterion. The first two phases of the aforementioned process define targets wishing to be achieved by the selection, attributes (criteria) as a base to perform alternative worthiness, determining the importance of the attributes, and a set of available alternatives for selecting the best one. Afterward, the decision maker evaluates available alternatives compared to the selected attributes. Regression analysis is one of the most exploited statistical methods that has great
application in the economy and other humanities. To explain, regression as a statistical method [5,6] allows estimations and assessment of a phenomenon based on the value of some other phenomenon or group of phenomena. Therefore x1, x2, ..., xk are undependable variables influencing and conditioning the size worthiness of dependable variable y. In experimental theory, undependable variables are called factors, while dependable variables are the results of the experiment. In regression analysis, there is the possible use of such units for determining performances, which are using one input or providing one output.

The presence of irrelevant and redundant attributes harms supervised learning performances; hence, the optimum set of attributes for learning makes fairly relevant non-redundant attributes and strongly relevant attributes. Due to the necessity of attributes selection method analysis, the subject of the research in this paper is the possibility of applying regression analysis as a classical statistical technique and AHP as a method of operational research in the process of supervised learning using the ReliefF classificatory modern ranking method. The method of previous learning shall determine sub-set of attributes based on accuracy estimation by applying decision tree classifiers after applying wrapper methods. WEKA [7] data mining tools were used for the application of selected classifiers, including intensive pattern discovery, which has been frequently studied during the last decade.

Evaluation of the proposed methodology was conducted using a case study predicting the choice of one of three possible modules for further study based on the assessment of subjects from the joint first year of study. The information on grades of the first-year graduate students at the Academy of Criminalistics and Police Studies in Belgrade was used as a dataset in this paper. Modules are presented as the attributes set. Analysis of the aforementioned sets should determine the prediction of the further orientation of students after the first year of studies. For each orientation, the meaningfulness of given attributes was determined by the application of traditional and smart methods to achieve an optimal sub-set that is relevant for further students' orientation. We use two traditional methods in our proposed model-AHP and regression analysis because, in general, we propose aggregation as methodology and because we first aggregate these two methods, which, as it is known, belong to two different types of methodology, i.e., subjective and objective. The comparison of accuracy prediction estimation is presented in the paper. It was performed by application of decision tree classifiers using wrapper methods and a set of attributes acquired by the application of ReliefF classifiers. Attributes whose combination acquires a satisfying prediction model were abstracted by the method of negative elimination; thus, these models were represented as relevant attributes present in both sets. Thereby, this paper presents an algorithm for extracting a minimal set of attributes that represent relevant data for further orientation in this case study. We propose novel label and unlabeled approaches for adaptively mining closed-rooted trees from data streams that change over time. It should be noted that closed patterns are a powerful type of frequent patterns due to the fact that they eliminate unnecessary information.

1.2. Related Studies

The use of statistical regression methods to predict student achievement can be found in the literature based on their success in graduation or entrance exam [8,9]. Moreover, statistical methods are used to determine the parameters for assessing whether a student will withdraw from studies or continue their studies based on the first results [10]. Regression models for predicting student academic performance in an engineering dynamics course are the considering subject of the solution with the same name [11]. Another paper [12] deals with academic progress using models based on linear and logistic regression, employing prior success and demographic factors as predictors.
The multi-criteria AHP method is used in higher education for the selection of candidates for teaching positions [13] where the AHP method is used to determine joint action as well as the priority of individual selection criteria. The selection of doctoral studies, depending on the goals that the applicant wants in his career, was done with the help of AHP [14]. Perspectives are set as a pseudo level in the hierarchy and for each of them, doctoral programs are offered and selected by the candidates depending on their preferences. The paper [15] gives a knowledge management system that recommends an orientation program to students who should choose it, and therefore, helps him be the most efficient in further studies. In the paper [16], comparative analyzes of DEA and AHP were performed, as well as their aggregation according to the importance of the first-year study subjects for selecting an appropriate study program on the example of the Faculty of Organizational Sciences in Belgrade. Significance of the weights of individual subjects was used to rank the benefits of study programs for a particular student, based on the success achieved in subjects that are common to all students in the first year of study. Moreover, very interesting papers are [17,18], in which the effects of the first-year course on student retention and the prediction of student success based on the achieved results are considered, respectively. The time dependence of predictions and ANOVA implementation in fuzzy AHP is discussed in [19].

Machine learning algorithms in [20,21] and data mining techniques in [22,23] are considered as one overview of student performance prediction modeling for further education in both pairs of these papers. Application of machine learning in predicting performance for computer engineering students is subject of the manuscripts [24,25] and data mining techniques are applied in predicting further education for students that study medical curriculum [26]. Additionally, supervised learning application is proposed in one model for student performance prediction [27]. Moreover, one comprehensive consideration of the application of supervised machine learning techniques was conducted in the prediction of students’ performance using different parameters including demographics and social interest [28].

A machine learning approach that uses two techniques, logistic regressions and decision trees, to predict student dropout at the Karlsruhe Institute of Technology is considered in [29]. We could also find such approach in [30], where a decision tree algorithm combined with the linear regression of data classification for students’ evaluation purposes is applied in Turkey.

2. Materials and Methods

These two subchapters are written to allow readers to replicate and build on published results. The proposed model is described in detail in the subchapter Methods, while a detailed description of the data set used for its development and evaluation is described before in the following subchapter Materials.

2.1. Materials

The dataset used in this case study for the evaluation of the proposed model in this paper and used data are taken from the students’ service database at the Academy of Criminalistics and Police Studies-CPA in Belgrade from 2006 to 2016. The analysis encompassed several graduated students in three departments with mutual modules in year one. Out of 290 graduated students at the Academy, 83 graduated from the Department of Safety (which is labeled as SD), 114 students graduated from the Criminology Department (labeled as CD), and 93 graduated from the Police Department (labeled as PD). For each student, we have compiled information on grades from 10 modules (i.e., subjects) in year one. These numbers are presented, along with their labels, in the following Table 1.
Table 1. Common subjects of students at year one.

| Subjects                        | Label |
|---------------------------------|-------|
| Fundamentals of economics       | $S_1$ |
| English language                | $S_2$ |
| Criminal law-general part       | $S_3$ |
| Criminal law-special part       | $S_4$ |
| Introduction to the Law         | $S_5$ |
| Special physical education      | $S_6$ |
| Sociology                       | $S_7$ |
| Police equipment                | $S_8$ |
| Constitutional Law              | $S_9$ |
| IT                              | $S_{10}$ |

As can be seen, descriptive statistics with the average grade of the students at the CPA in the above modules at year one, observed for each department, are presented. For these purposes, the group of statistical methods, based on the methods of descriptive statistical analysis of the observable data, were used. The values of minima (MIN), maxima (MAX), average grade (MEAN), median grade (MEDIAN), mode grade (MODE), standard deviation (STDEV), sample variance (VAR), kurtosis (KURT) and skewness (SKEW) are set in rows of the following Tables 2–4.

As can be seen from the presented tables, the minimum and maximum grades of students in all subjects are equal to MIN = 6 and MAX = 10. The next three rows show the so-called dominant characteristics (MEAN, MEDIAN and MODE). Obviously, the students showed the best success in the subject ‘English language’ (module $S_2$), while other subjects have lower values of these statistical indicators. In the next two rows, the values of the so-called variability measure, standard deviations (STDEV) and variance (VAR) are given. In all three sets of observed data, i.e., for all three departments above, slightly statistically-significant differences in the variability of the observed subjects can be seen.

Table 2. Descriptive statistics of the success of students at Department of Safety (sample size is $N = 83$)

| Items   | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| MIN     | 6     | 6     | 6     | 6     | 6     | 6     | 6     | 6     | 6     | 6       |
| MAX     | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10      |
| MEAN    | 7.29  | 8.17  | 7.12  | 7.24  | 7.41  | 7.76  | 7.02  | 7.51  | 7.41  | 7.37    |
| MEDIAN  | 7     | 8     | 7     | 7     | 8     | 7     | 7     | 7     | 7     | 7       |
| MODE    | 6     | 8     | 7     | 6     | 7     | 7     | 6     | 6     | 7     | 6       |
| STDEV   | 1.365 | 1.093 | 1.029 | 1.241 | 1.204 | 1.280 | 1.107 | 1.325 | 1.264 | 1.428   |
| VAR     | 1.862 | 1.195 | 1.060 | 1.539 | 1.449 | 1.639 | 1.224 | 1.756 | 1.599 | 2.038   |
| KURT    | −0.695| −0.876| 0.407 | −0.285| −0.402| −0.728| 0.633 | −0.916| −0.405| −0.908  |
| SKEW    | 0.740 | 0.248 | 0.901 | 0.748 | 0.573 | 0.408 | 0.997 | 0.447 | 0.709 | 0.664   |

Table 3. Descriptive statistics of the success of students at Criminology Department (sample size is $N = 114$)

| Items   | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| MIN     | 6     | 6     | 6     | 6     | 6     | 6     | 6     | 6     | 6     | 6       |
| MAX     | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10    | 10      |
The last two rows of Tables 2–4 show the values of kurtosis, i.e., the elongation coefficient (KURT), as well as the skewness coefficient (SKEW). One can notice that most of the observed data (except, for instance, subject S5 in the Police Department) have negative values (KURT < 0). This means that the observed data sets have the so-called platykurtic properties, which means that they are 'flatter' compared to a normal distribution. Finally, the obtained values of the skewness coefficient are, in most cases, positive (SKEW > 0), which means a higher frequency (and consequently mode values) of lower student grades is present.

2.2. Methods

The model, which is proposed in this paper, uses a traditional method of regression analysis which is first aggregated with the AHP method and, in this way, the obtained method has been combined using the application of the ReliefF classificatory ranker method of machine learning and the decision trees classifiers J48 has been used for the aggregation purpose. All of these methods and the proposed aggregated model are studied in detail in four subsections of this subchapter.

2.2.1. Traditional Techniques

The problem of prediction in the choice of study direction is observed in this paper, from the point of view of the mathematical approach. The solution to the problem is to determine the impact of individual subjects in the first year as criteria that affect the final success of the student, which is the observed goal and later allows for the creation of a model that is based on the success of individual subjects in the first year and can predict the choice of appropriate study. Practically, this problem, when it comes to the traditional mathematical and statistical approach, is reduced to the well-known in the literature problem of determining the weights of criteria in a multicriteria problem and a linear regression model.

The classification of mathematical methods for determining the weights of criteria is not uniform. Namely, the division between these methods is made following the authors’
concept and the need to solve a practical problem. There are different divisions of methods, such as: statistical and algebraic, direct and indirect, holistic and decomposed, and compensatory and non-compensatory, which can be seen more in [31].

The most important groups of objective and subjective methodologies for determining the weight of the criteria are discussed in [32], as follows:

- Standard methods of statistical analysis, including the most commonly used regression and correlation analysis as well as variance and factor analysis. These methods require a statistically relevant sample as well as meeting the strict condition of the relationship between the number of members of the observed sample and the number of criteria and imply a normal distribution, which is eliminated by using so-called non-parametric statistical methods;

- Methods of operational research, most often methods of multicriteria analysis, which are mostly algebraic but also statistically-based methods for determining the weight of the criteria and whose main task is in fact to choose the optimal number of alternatives and optimal solutions. Methods of operational research that are less demanding than statistical, as well as non-parametric statistical methods, are used where the condition of normal distribution in the sample is not met as well as the required sample size and its relationship with the number of considered criteria. When it is necessary to choose alternatives and their ranking and quantitative positioning;

Methods of operational research can be divided into two basic groups [32]: subjective methods, which depend on the influence of decision makers on the difficulty of criteria and objective methods, which are based on the application of objective mathematical, primarily statistical, apparatus to the information contained in the matrix decision making [31,33].

Objective methods from the essential information of each of the evaluated criteria determine their weight according to two principles [34–36]:

- The criterium that has the least variation in the considered cases, i.e., alternatives have the least impact, i.e., weight; this is the principle of contrast intensity within when special methods appear as those in which entropy, standard deviation, variation, etc. are most often used as a measure of contrast intensity;

- The criterium that is in conflict with a large number of others is more important, and that is the principle of conflict character in which the most well-known method of correlation is the correlation of the coefficient between pairs of criteria.

In the literature, one can find the application of statistical regression methods to predict students’ success [16] based on their success in graduation or the entrance exam [8,19]. Moreover, statistical methods are used to determine the parameters for assessing whether a student will withdraw from their studies or continue their studies based on the first results [10].

Supervised learning is considered through its main aim to obtain the best prediction results in [37,38] and as one overview in [39]. In [24,40], the applications of machine learning in predicting performance for computer engineering students and disease prediction were considered, respectively.

Regression Analysis

Regression analysis is a method for examining the influence of more different independent variables on one dependent variable to determine the analytic form of this connection, i.e., the model which will be used in analytical and predictive applications. It has a deterministic model in which each value of independent variables exists exactly one value of the dependent variable and which can be given in the form:

\[
y = a + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_px_p
\]

(1)

where \(b_i\) are partial coefficients of regression.
In the case when we have a non-explained variation of the dependent variable in the form of error which we notate as a consequence of the influence of random action or non-including all existing independent variables, we can present multiple regression in the form:

\[ y = a + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_px_p + e_i \]  

The calculation of parameters \(a\) and \(b\) in Equation (2) can be executed with the method of least squares with minimization of the sum of squares of the residuals, and in a general case, notated as:

\[ \sum_{i=0}^{n}(y_i - a - b_1x_{i1} - b_2x_{i2} - \Lambda - b_px_{pi})^2 \]  

In practice, an algebraic algorithm for solving arising system of equations is rarely in use compared with the well-known Gaussian method of multiplication or with the analysis of variance-ANOVA that is mostly used in empiric research.

One of the particularly important problems that may occur in the multiple regression model is a mult-correlation. It manifests as an intercorrelation between explanatory variables and usually leads to erroneous results of regression analysis [41, 42]. In order to check the multi-correlation in the regression model that we obtained here, two diagnostic tools were used and described in detail below (please see Section 3 ‘Results and Findings’). Firstly, the correlation coefficients between all predictor variables were computed, and the weak or moderate correlation was noticeable. In addition, for the purpose of double-checking, another measure of multi-correlation, a well-known variance inflation factor (VIF), was also used.

Multi-Criteria Decision Making

Multi-criteria decision making is a field that is receiving great importance in the last two decades since each process requires contemplation of numerous criteria that are often in conflict or stated in different measurement units. Since the 1960s, further on, a various number of multi-criteria analyses have been developed and could be classified on several grounds. One of the most important multi-criteria classification methods of decision making was conducted by Hwang and Yoon [43], who had classified 17 different methods per type and relevant features of information brought by the decision makers.

According to the type of information, all stated methods are divided into two groups:

1. Methods without information on attributes,
2. Methods requesting certain attribute information.

There are several ways to enable attribute transformation and adjust them to the models of multi-criteria decision making: conversion of the attributes to the scale interval, normalization of the attributes and assembling of the appropriate weight set. Another way of transformation is the assignment of an appropriate weight set, and it is mostly used in situations when it is necessary to determine the relevant importance of certain attributes. For the \(N\) criterion, the weight set is:

\[ t^T = (t_1, t_2, \cdots, t_j, \cdots t_n) \]  

\[ \sum_{j=0}^{n} t_j = 1 \]  

Among numerous estimation techniques of the relative importance of certain abstracted attributes are methods of natural vectors, method of weight in smallest quadrants, entropy method, etc. The algorithm of the AHP method can be described as a structural analysis of one complex decision-making problem that contains more criteria, more alternatives, even to have more decision makers (decision-making group), determining relevant weight criteria and alternatives per level and forming a final alternative outcome (rang alternative). This process is one of the well-known methods of multi-criteria decision making, which is mostly used in cases when there is the possibility
of a hierarchical structure of relevant criteria. This method was created in the 1970s by Thomas Saaty [44].

It is important to note that, since the main part of the AHP method is pairs’ comparison of hierarchy elements and formation of appropriate local reciprocal numerical matrices, from which the weights of compared elements are determined, these matrices with the calculated element’s weights carry the measurable information on consistency used by the consistency ratio (CR). A model that provides CR <0.1 is considered a good one.

2.2.2. Knowledgeable Data Analysis

The basic property differing deep data analysis from the traditional or ‘ordinary’ analysis is the application of machine learning. Intelligent data analysis is another term for deep data analysis. Adjective ‘intelligent’ emphasizes that this data analysis is based on artificial intelligence proceedings. Attribute selection is the field that was developed within the framework of sample recognition in mathematical statistics [45], knowledge revelation [46], machine learning [47,48], especially neuron grids, and many other fields. The fundamental task of the attributes’ selection is the reduction of spatial dimensionality and removal of redundant, irrelevant and disrupted data that speeds up the operation of the learning algorithm, improves the data quality and increases the accuracy of taught knowledge. When selecting attributes, there are two approaches that are different in what they need to achieve. The first approach is finding and ranking the sub-set of the attributes useful for the construction of the quality model. Another approach is finding or ranking all potentially important attributes. Using a sub-set of potentially important attributes for model construction is presented as a sub-optimal approach with the construction of some models [49]. Furthermore, the issue of usefulness versus the importance of the attributes could be very interesting [50].

Attribute selection Using ReliefF Algorithm

The complexity of group correlation analysis is a result of a huge number of combinations of the attributes, which relations should be considered as \( O(2^N) \), where \( N \) is the number of attributes in the model [51]. Due to the usual great complexity, the approximation is applied. For example, only partial analysis of individual attributes with the class \( O(N) \), or analysis of only some of the possible combinations (interaction of length 2 or 3 attributes) is performed. Entropy is a commonly used measure in information theory [52], which characterizes the purity of an arbitrary collection of examples. The entropy measure is considered a measure of the system’s unpredictability. The entropy of \( Y \) is:

\[
H(Y) = -\sum_{y \in Y} p(y) log_2(p(y))
\]

where \( p(y) \) is the marginal probability density function for the random variable \( Y \). If the observed values of \( Y \) in the training data set \( S \) are partitioned according to the values of a second feature \( X \), and the entropy of \( Y \) concerning the partitions induced by \( X \) is less than the entropy of \( Y \) before partitioning, then there is a relationship between features \( Y \) and \( X \). The entropy of \( Y \) after observing \( X \) is then:

\[
H \left( Y \mid X \right) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y \mid x) log_2 p(y \mid x)
\]

where \( p(y \mid x) \) is the conditional probability of \( y \) for given \( x \).

ReliefF is a filtering method with proceedings of attribute ranking. It is based on the procedure of the k-nearest neighbors (k-NN). Figure 1 shows the ReliefF algorithm. This algorithm estimates and ranks each attribute with the global grade function \([-1, \ldots, 1]\). Weight calculation could be performed based on the probability of the nearest neighbors from two different classes with different values of the attributes as well as on the probability of two neighbors from the same class having the same value of the attributes.
The function \( \text{diff} \) \((\text{Attribute}; \text{Instance1}; \text{Instance2})\) calculates the difference between the values of the attribute for two instances.

```
set all weights \( W[A] \rightarrow 0.0 \)
for \( i = 1 \) to \( m \) do
    begin
        randomly select an instance \( R \)
        find \( k \) nearest hits \( H_j \)
        for each class \( C=\text{class}(R) \) do
            find \( k \) nearest misses \( M_j(C) \)
            for \( A = 1 \) to \( \text{attributes} \) do
                \[
                W[A] - W[A] - \frac{1}{m} \sum_{j=1}^{k} \text{diff}(A, R, H_j) \cdot \frac{1}{m-k} \sum_{j=1}^{k} \text{misses}(A, R, M_j(C)) \cdot \frac{1}{m-k}
                \]
        end
```

**Figure 1.** ReliefF algorithm.

When considering discrete attributes, the difference is either 1 (when values are different) or 0 (when values are the same), while in continuous attributes case, the difference is an actual difference normalized to the interval \([0, 1]\). Kononenko notes in [53] that the higher value of \( m \) (the number of sampled instances), the more reliable the ReliefF’s estimates are. Of course, it should be noted that the increasing \( m \) rises the running time.

### 2.2.3. Classification Technique and Algorithms

Classification is the task of data mining that administers the separation of data set examples into previously determined classes of output variables based on the value of input variables [54]. Classification of some objects is based on finding similarities with priory-determined objects belonging to different classes, whereas the similarity of two objects is determined by analyzing their characteristics. The task is to design a model based on the characteristics of the objects with a previously known classification. That model shall represent the ground to perform the classification of new objects. In the problem of classification, several classes are known in advance and limited. By testing, the classifier performs classification of the test set examples in priory determined classes of attributes. If the classifier makes some errors in test data or if there is a higher percentage of wrongly classified examples, the conclusion is that a wrong and unstable model is created. In that case, it is necessary to perform improvements to the model by modifying the applied classification process. Up-to-date research shows that most applied classifiers include Bayes networks, decision trees, neural networks, support vector machines, and K–nearest neighbors [55].

Decision trees are well-known classification techniques considering that they include several ways of creation of easily interpreted trees used for the classification of categories and numerical values of the attributes. These classification methods perform the division of data to nodes and leaves until the entire set of data is being analyzed. The well-known algorithms are ID3 [56] and C4.5 [57]. Algorithm C4.5 for decision-making tree induction is developed on the basis of the ID3 algorithm, with multiple significant improvements compared to the basic algorithm: imperfect operation with continual attributes and emitted values of the attributes, new estimation of gain ration quality, and simplification of the taught tree due to the increasing classification accuracy of new examples. It is available as an independent program as the objective module (library MLC++) for usage within the other systems for supervised learning and intelligent data analysis. J48 algorithm of Weka software is a popular machine learning algorithm based upon Quilan’s C4.5 algorithm [58].
Classifiers Estimation

Classifiers estimation enables the prediction of performances, selection of the best and behavioral assessment of multiple different classifiers. Cross-validation is a statistical method of assessing and comparing learning algorithms by dividing data into two segments: one is used to learn or train a model, and the other one is used for model validation. In standard cross-validation, the training and validation sets must cross over in successive rounds such that each data point has a chance of being validated.

The basic form of cross-validation is k-fold cross-validation. When we talk about the k-fold cross-validation, the whole data is partitioned into k equally (or nearly equal) sized segments/folds. Afterward, k iterations of training and validation are completed in such a way that within each iteration, a different fold of the data is held-out for validation while the remaining k = 1 folds are used for learning. In most cases in data mining and machine learning, 10-fold cross-validation (k = 10) is the most common one. J48 classifier with the corresponding parameters is used and evaluated with this mostly used type of cross-validation.

The biggest number of measurements for classifying models’ estimation is related to classification issues with two classes. That does not represent a specific limitation for the application of those measurements, taking into consideration that the issues with the higher number of classes can be shown in the form of a sequence of problems with two classes. Each of them is separately abstracted for one of the classes as the target ones, a the data set is divided into a positive and negative example of the target class, but the negative ones are the examples of all other classes. Information Retrieval (IR) measures community has developed three measures in everything according to Figure 2.

In Figure 2 so called a 2 × 2 confusion matrix which enables calculation formulas is presented:

\[ TP + TN + FN + FP = N \]  

where N is the total number of members in the considered set which will be classified.

It is necessary to notice that these numbers are counts, i.e., integers, not ratios or fractions.

The accuracy, precision, recall and F1 measure can be respectively calculated as:

\[ \text{Accuracy} = \frac{TP + TN}{N} \]  

\[ \text{Precision} = \frac{TP}{TP + FP} \]  

\[ \text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \]  

\[ \text{F - measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]
An evaluation of the prediction performance of a classifier is carried out using the method called Receiver Operating Characteristic (ROC) curves, which represents the rate of false-positive cases and the rate of true-positive cases on the OX and the OY axes, respectively. ROC curves and parameter AUC (area under the curve) enable best-way evaluation for checking each classification model in Figure 3 [59].

In theory, it is known that for the values of AUC = 70%, it is said that the classification is good and sometimes leads to the highest possible values of 100%.

2.2.4. Proposed Model for Selection of the Relevant Attributes

In the phase of data preparation, in the data set for three departments, the Attribute Department was added, and its value was appointed as binominal. For the Attribute Department, the (yes, no) values were determined. Attribute Department is marked as class and response variable, respectively. Educational information is mostly transparent
since they do not contain wrong values collected automatically (log files, database with grades, database of learning system). As the data with the average grade of the year one module at the CPA on all three bases were in numerical form for each graduated student, it was not necessary to apply some of the standard discretization filters during the data preparation.

After the data were acquired and extracted into one unique direction selection, the data mining algorithm was applied as a method and used to estimate and select the attributes. The aim is to extract irrelevant and redundant attributes from the learning data set. In order to keep the data learning set with the desirable attributes only, the measure of attributes evaluation compared to the classification problem is necessary. Filtering methods encompass techniques for attribute values assessment relying on heuristics based on general features of data.

The proposed model selects relevant attributes by combining acquired aggregation results obtained using the ReliefF algorithm, shown in Figure 4.

A selection of the attributes using various filtering techniques that are performing ranking according to the importance estimation is carried out in the WEKA system so that the attributes are ranked and assessed using the current training set. For the importance assessment of the attributes, the ReliefF algorithm with the ranker method was used. Extraction of the relevant attribute set from the whole set was performed by the negative elimination process [60,61].

![Figure 4. Suggested model of relevant attributes selection.](image)

The weakest ranked attribute is extracted in the next step of attribute ranking. After each extraction, the worst-ranked attribute was selected in WEKA software, and the estimation accuracy that is unique for all steps within these proceedings for the selected learning method was chosen.

3. Results and Findings

For determination of the relevant attributes for selection of the study program at the CPA in Belgrade, the evaluation aggregation was used, as a combination of regression
analysis and subjective AHP method, respectively, in order to points gained grade and eliminate the bad side of both applied technologies. In order to calculate relevant attributes, the following procedure was used:

Step 1: In AHP analysis, the procedure has been started from the data from Tables 2–4 of the average grades per subject for each department individually, safety, criminology and police, respectively. They are compared in pairs. The result of the comparison by pairs is the formation of a reciprocal matrix in accordance with the preferences defined by the Saaty scale. Saaty scale is a liqueur type and contains divisions from 1 to 9, where 1 is the same meaning of the criteria, and 9 is the absolute meaning of one in relation to the other (even numbers are intermediate divisions on the scale). Moreover, the scale contains reciprocal values, for example if p1: p2 = 5, then P2: P1 = 1/5. What importance will be given to a certain ratio of average grades depends on the subjective opinion of the decision maker. Logically, the highest possible grade is 10, and the lowest positive grade is 6. Hence, the ratio of the significance of subjects whose grades are P1 = 10 and P2 = 6 can be determined as 10:6 = 1.66667, and, on the Saaty scale, it is the absolute significance of P1 in relation to P2 and enter 9. Conversely, the ratio P2: P1 is the lowest possible preference 6:10 = 0.6. Therefore, it is necessary to distribute preferences in the interval from 0.6 to 1.66667, where the same meaning of the criterion must be marked with 1. Precisely, because it is necessary to distinguish between similar estimates (approximate values), the divisions are distributed so that the final values occupy a larger interval. Values from 0.88 to 1.10 are divided into several smaller intervals. In that way, reciprocal matrices were formed for each direction, and further procedure of applying the AHP method involved summing each of the columns of the reciprocal matrix, so the next matrices were constituted, which were normalized values of the reciprocal matrix (each coefficient divided by the sum of the corresponding column). In the end, all normalized coefficients are added into rows and the weighting coefficient is obtained as the quotient of the sum of the rows, and the number of criteria used in the model.

AHP analysis was implemented by using VCO (multi-criteria decision-making software package of the Faculty of Organizational Sciences in Belgrade) using hierarchy and AHP analysis model [62]. Obtained results are shown in Table 5.

Table 5. AHP analysis.

| Subjects Average Grade of | Department of SAFETY | Criminology Department | Police Department |
|--------------------------|----------------------|------------------------|------------------|
| S1                       | 0.0781               | 0.0576                 | 0.0356           |
| S2                       | 0.3378               | 0.3116                 | 0.3048           |
| S3                       | 0.0303               | 0.0367                 | 0.0349           |
| S4                       | 0.0402               | 0.0255                 | 0.0461           |
| S5                       | 0.0796               | 0.1026                 | 0.1087           |
| S6                       | 0.188                | 0.1026                 | 0.0616           |
| S7                       | 0.0222               | 0.1026                 | 0.0832           |
| S8                       | 0.0904               | 0.1026                 | 0.1087           |
| S9                       | 0.0713               | 0.1003                 | 0.1785           |
| S10                      | 0.0617               | 0.0576                 | 0.0374           |

Reciprocal AHP numerical matrices for all departments provide CR < 0.1, so the model can be considered as a good one.

Step 2: Regression analysis was implemented by using model (1) of linear regression in the SPSS software package [63,64], and the results are shown in Table 6. One can notice that the last row contains the estimated values of the multiple coefficients of determination ($R^2$). It represents a widely used quantitative measure of agreement of
the theoretical, fitting model in relation to the given set of empirical data. The estimated values of $R^2$ confirm that the obtained regression model can be an adequate model of dependency for the given subjects $S_1, \ldots, S_{10}$.

Table 6. Regression analysis model with standardized Beta-coefficients.

| Model | Department of Safety | Criminology Department | Police Department |
|-------|-----------------------|------------------------|-------------------|
| (Constant) | Beta | Beta | Beta |
| $S_1$ | 0.104 | 0.061 | 0.084 |
| $S_2$ | 0.024 | 0.191 | 0.01 |
| $S_3$ | 0.133 | 0.119 | 0.205 |
| $S_4$ | 0.144 | 0.034 | 0.21 |
| $S_5$ | 0.129 | 0.023 | 0.111 |
| $S_6$ | 0.025 | 0.085 | 0.109 |
| $S_7$ | 0.009 | 0.066 | 0.07 |
| $S_8$ | 0.126 | 0.154 | 0.077 |
| $S_9$ | 0.079 | 0.133 | 0.087 |
| $S_{10}$ | 0.228 | 0.179 | 0.037 |
| $R^2$ | 0.840 | 0.895 | 0.836 |

Legend: Abbreviations in Table 6 correspond to the ones in Table 1.

As already pointed out in the previous section, one of the most important problems that arise here is the potential multi-correlation of the predictors in the above regression models. In order to determine the intercorrelation relationship between the predictors $S_i$ and $S_j$, where $i,j = 1, \ldots, 10$, the following Tables 7–9 show the estimated values of their correlation coefficients ($R_{ij}$).

Table 7. Correlogram of the grades of students between different subjects at the Department of Safety.

| Subjects | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| $S_1$ | 1.000 | 0.217 | 0.365 | 0.459 | 0.244 | -0.015 | 0.061 | 0.385 | 0.319 | 0.367 |
| $S_2$ | 0.217 | 1.000 | 0.314 | 0.393 | 0.211 | -0.041 | 0.348 | 0.352 | 0.291 | 0.343 |
| $S_3$ | 0.365 | 0.314 | 1.000 | 0.563 | 0.604 | 0.327 | 0.217 | 0.576 | 0.690 | 0.054 |
| $S_4$ | 0.459 | 0.393 | 0.563 | 1.000 | 0.567 | 0.290 | 0.269 | 0.652 | 0.671 | 0.287 |
| $S_5$ | 0.244 | 0.211 | 0.604 | 0.567 | 1.000 | 0.327 | 0.124 | 0.443 | 0.579 | 0.011 |
| $S_6$ | -0.015 | -0.041 | 0.327 | 0.290 | 0.327 | 1.000 | 0.234 | 0.311 | 0.265 | -0.292 |
| $S_7$ | 0.061 | 0.348 | 0.217 | 0.269 | 0.124 | 0.234 | 1.000 | 0.298 | 0.118 | 0.247 |
| $S_8$ | 0.385 | 0.352 | 0.576 | 0.652 | 0.443 | 0.311 | 0.298 | 1.000 | 0.586 | 0.321 |
| $S_9$ | 0.319 | 0.291 | 0.690 | 0.671 | 0.579 | 0.265 | 0.118 | 0.586 | 1.000 | 0.246 |
| $S_{10}$ | 0.367 | 0.343 | 0.054 | 0.287 | 0.011 | -0.292 | 0.247 | 0.321 | 0.246 | 1.000 |

Table 8. Correlogram of the grades of students between different subjects at the Criminology Department.

| Subjects | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------|
| $S_1$ | 1.000 | 0.312 | 0.428 | 0.497 | 0.457 | 0.394 | 0.327 | 0.511 | 0.423 | 0.300 |
| $S_2$ | 0.312 | 1.000 | 0.193 | 0.341 | 0.375 | 0.292 | 0.386 | 0.216 | 0.281 | 0.118 |
| $S_3$ | 0.428 | 0.193 | 1.000 | 0.582 | 0.495 | 0.259 | 0.434 | 0.536 | 0.365 | 0.419 |
| $S_4$ | 0.497 | 0.341 | 0.582 | 1.000 | 0.464 | 0.344 | 0.545 | 0.488 | 0.409 | 0.537 |
| $S_5$ | 0.457 | 0.375 | 0.495 | 0.464 | 1.000 | 0.263 | 0.442 | 0.597 | 0.347 | 0.219 |
| $S_6$ | 0.394 | 0.292 | 0.259 | 0.344 | 0.263 | 1.000 | 0.165 | 0.276 | 0.224 | 0.177 |
Table 9. Correlogram of the grades of students between different subjects at the Police Department.

| Subjects | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| $S_1$    | 1.000 | -0.175| 0.174 | 0.194 | 0.476 | 0.060 | 0.124 | 0.161 | 0.282 | 0.319  |
| $S_2$    | -0.175| 1.000 | 0.366 | 0.073 | 0.002 | 0.153 | 0.078 | 0.117 | 0.062 | -0.134 |
| $S_3$    | 0.174 | 0.366 | 1.000 | 0.599 | 0.428 | 0.139 | 0.285 | 0.168 | 0.425 | 0.331  |
| $S_4$    | 0.194 | 0.073 | 0.599 | 1.000 | 0.364 | 0.289 | 0.265 | 0.224 | 0.245 | 0.333  |
| $S_5$    | 0.476 | 0.002 | 0.428 | 0.364 | 1.000 | 0.353 | 0.249 | 0.362 | 0.267 | 0.515  |
| $S_6$    | 0.060 | 0.153 | 0.139 | 0.289 | 0.353 | 1.000 | 0.239 | 0.066 | 0.000 | -0.050 |
| $S_7$    | 0.124 | 0.078 | 0.285 | 0.265 | 0.249 | 0.239 | 1.000 | 0.120 | 0.344 | 0.265  |
| $S_8$    | 0.161 | 0.117 | 0.168 | 0.224 | 0.362 | 0.066 | 0.120 | 1.000 | 0.344 | 0.296  |
| $S_9$    | 0.282 | 0.062 | 0.425 | 0.245 | 0.267 | 0.000 | 0.344 | 0.344 | 1.000 | 0.242  |
| $S_{10}$ | 0.319 | -0.134| 0.331 | 0.333 | 0.515 | -0.050| 0.265 | 0.296 | 0.242 | 1.000  |

As it can be seen, students’ grades at the Department of Safety satisfy the condition $\max R_{ij} < 0.7$, when $i \neq j$. On the other hand, in the case of the other two departments, the inequality $\max R_{ij} < 0.6$ holds. Thus, there is only a weak or (in the ‘worst case’) moderate correlation between students’ grades in different subjects in all three departments.

Multicollinearity can be expressed in another way, using the so-called variance inflation factor (VIF), which represents the measure of the severity of multicollinearity in regression analysis. Mathematically, for a multiple regression model given by Equation (1), the VIF can be expressed using the coefficient of determination ($R^2_i$) of the new multiple regression model with one explanatory variable ($x_i$) as the response variable and the other variables $x_j$ ($j \neq i$) as its explanatory variables:

$$VIF = \frac{1}{1 - R^2_i}$$  \hfill (12)

The calculated VIF values of each of the three previously obtained regression models are shown in the following Table 10.

Table 10. Estimated values of the coefficient of determination ($R^2_i$) and the variation inflated factor (VIF) for all modules, i.e., subjects ($S_i$, $S_2$, ..., $S_{10}$).

| Department        | Items  | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ | $S_9$ | $S_{10}$ |
|-------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| Safety            | VIF    | 1.541 | 1.429 | 2.519 | 2.899 | 3.003 | 1.592 | 1.414 | 2.227 | 3.846 | 1.934  |
| Criminology       | VIF    | 0.409 | 0.280 | 0.490 | 0.595 | 0.746 | 0.217 | 0.443 | 0.493 | 0.688 | 0.359  |
| Police            | VIF    | 1.692 | 1.389 | 1.961 | 2.469 | 3.937 | 1.277 | 1.795 | 1.972 | 3.205 | 1.560  |
| Department        | $R^2_i$| 0.299 | 0.309 | 0.613 | 0.470 | 0.572 | 0.330 | 0.228 | 0.276 | 0.361 | 0.418  |
|                   | VIF    | 1.427 | 1.447 | 2.584 | 1.887 | 2.336 | 1.493 | 1.295 | 1.381 | 1.565 | 1.718  |

It can be seen that for all these three sets of data, i.e., the grades of students from different departments, the condition $\max(VIF) < 4$ is valid. This result is fully consistent with both practical and theoretical interpretations of this coefficient [65,66]. In other words, the values of VIF, obtained in this way, indicate that the multi-collinearity between the predictors in the previously obtained regression models is not emphasized and, therefore, can be neglected.
Step 3: Evaluation aggregation. Aggregated evaluation measures are acquired by the simple application of arithmetic middle of acquired AHP results and regression analysis in order to objectify the acquired grade and eliminate the bad sides of both applied methodologies. The results are shown in Table 11. So as to clarify, the evaluation aggregation set is given as follows:

\[ \text{aggr}_{ik} = \frac{(\text{regression\_analysis}_{ik} + \text{AHP}_{ik})}{2} \]  

(13)

Table 11. Evaluation aggregation.

| First year Modules | Department for Safety | Criminology Department | Police Department |
|--------------------|------------------------|------------------------|------------------|
|                    | Regr. | AHP | Aggr. | Regr. | AHP | Aggr. | Regr. | AHP | Aggr. |
| S₁                 | 0.105 | 0.0781 | 0.09155 | 0.059 | 0.0576 | 0.0583 | 0.084 | 0.0356 | 0.0598 |
| S₂                 | 0.024 | 0.3378 | 0.1809 | 0.183 | 0.3116 | 0.2473 | 0.010 | 0.3048 | 0.1574 |
| S₃                 | 0.134 | 0.0303 | 0.08215 | 0.113 | 0.0367 | 0.07485 | 0.204 | 0.0349 | 0.11945 |
| S₄                 | 0.143 | 0.0402 | 0.0916 | 0.034 | 0.0255 | 0.02975 | 0.210 | 0.0461 | 0.12805 |
| S₅                 | 0.129 | 0.0796 | 0.1043 | 0.021 | 0.1026 | 0.0618 | 0.110 | 0.1087 | 0.10935 |
| S₆                 | 0.026 | 0.188 | 0.107 | 0.082 | 0.1026 | 0.0923 | 0.111 | 0.0616 | 0.0885 |
| S₇                 | 0.008 | 0.0222 | 0.0151 | 0.063 | 0.1026 | 0.0828 | 0.071 | 0.0832 | 0.0771 |
| S₈                 | 0.125 | 0.0904 | 0.1077 | 0.147 | 0.1026 | 0.1248 | 0.077 | 0.1087 | 0.09285 |
| S₉                 | 0.078 | 0.0713 | 0.07465 | 0.127 | 0.1003 | 0.11365 | 0.087 | 0.1785 | 0.13275 |
| S₁₀                | 0.228 | 0.0617 | 0.14485 | 0.171 | 0.0576 | 0.1143 | 0.037 | 0.0374 | 0.0372 |

Step 4: According to the proposed model of aggregation of traditional methods of regression and AHP with the method of classification from the group of machine learning algorithms which is given in previous Section 2.2.4, we apply the ReliefF algorithm with the ranker method for importance assessment of the attributes. After that, we apply accuracy evaluation for both groups of methods using decision trees classifiers J48 and, for relevant attributes, we take the number of the most important attributes, for which the deviation of accuracy for these two groups of methods is the smallest [67–70]. The results for the Department of Safety on CPA are given in Tables 12–14.

Table 12. Ranked attributes by ReliefF on the Department for Safety dataset.

| Ranked Attributes | Attributes |
|-------------------|------------|
| 0.1862            | S8         |
| 0.1279            | S5         |
| 0.1268            | S1         |
| 0.1198            | S4         |
| 0.1163            | S9         |
| 0.0955            | S10        |
| 0.0766            | S2         |
| 0.0625            | S6         |
| 0.0466            | S3         |
| 0.0455            | S7         |

Table 13. Accuracy of the Department for Safety dataset with J48 tree.

| Weighted Avg. | No. Attributes | Precision | Recall | F-Measure | AUC |
|---------------|----------------|-----------|--------|-----------|-----|
|               | 10             | 0.854     | 0.854  | 0.854     | 0.860 |
Step 5: By aggregation evaluation for the same department, we have acquired the following measurement values shown in Table 14.

| No. | Attribute | Precision | Recall | F-Measure | AUC |
|-----|-----------|-----------|--------|-----------|-----|
| 10  |           | 0.854     | 0.854  | 0.854     | 0.795|
| 9   |           | 0.830     | 0.854  | 0.854     | 0.795|
| 8   |           | 0.854     | 0.854  | 0.854     | 0.795|
| 7   |           | 0.854     | 0.854  | 0.854     | 0.795|
| 6   |           | 0.794     | 0.805  | 0.797     | 0.790|
| 5   |           | 0.727     | 0.756  | 0.733     | 0.669|
| 4   |           | 0.876     | 0.878  | 0.870     | 0.794|
| 3   |           | 0.819     | 0.829  | 0.818     | 0.735|
| 2   |           | 0.572     | 0.756  | 0.651     | 0.485|
| 1   |           | 0.572     | 0.756  | 0.651     | 0.485|

The number of the important factors should be 4 because the deviation of the accuracy of the Department for Safety dataset with ranked attributes by traditional aggregation and with classification is \(\epsilon = 0.854 - 0.876 = 0.022\), and it is less than when 3 of them are taken as important factors because then it is \(\epsilon = 0.883 - 0.819 = 0.064\).

**4. Discussion and Suggestions**

By applying attribute rank gained by aggregation, it could be noticed that the set of 4 attributes gave the highest mid-estimation values for the Department of Safety. Attributes presented as relevant on the basis of this criterion are: \(S2, S10, S8\) and \(S6\). Acquired mid-accuracy values with a set of attributes that was decreased by negative elimination of the lowest ranked by applying both of the techniques for the Department of Safety are shown in Figure 5.
Figure 5. Attributes ranked by aggregation of the Department of Safety dataset.

Presented accuracy results for both cases show that these proceedings enable the discovery of the approximate accuracy values on the grounds, which we could determine an equal number of attributes in both sets. The comparison was performed in order to achieve the aforementioned, i.e., differences of set accuracy of ranking attributes by aggregation technique PrecA(k) and set of ranked attributes by applying ReliefF algorithm PrecR(k). In consideration of accuracy precision $\varepsilon$, the values of sets with sequences of 1 to 10 values were not taken into consideration. On the basis of acquired results (Figure 5), the accuracy precision is visible by applying both ranking methods, the smaller with the set of 4 attributes. For the other two departments, the closest accuracy by application of both methods was acquired by the set of 8 attributes for the Criminology Department and the set of 9 attributes for the Police Department, shown in Figure 6 and Figure 7, respectively.
Figure 7. Attributes ranked by aggregation of the dataset of Police Department.

Set of the appropriate number of attributes with ranking coefficient that we have acquired on the basis of previous figures for all Departments is shown in the Table 15.

Table 15. Number of attributes with ranking coefficient for all Departments.

| No. Attr. | Department of Safety | Criminology Department | Police Department |
|-----------|----------------------|------------------------|------------------|
|           | Relieff              | Aggreg.                | Relieff          | Aggreg.          | Relieff          | Aggreg.          |
| 1         | S8                   | S2                     | S5               | S2               | S1               | S2               |
| 2         | S5                   | S10                    | S2               | S8               | S9               | S9               |
| 3         | S1                   | S8                     | S3               | S10              | S5               | S4               |
| 4         | S4                   | S6                     | S9               | S9               | S3               | S3               |
| 5         |                     | S4                     | S6               | S7               | S7               | S5               |
| 6         | S8                   | S7                     | S8               | S8               |                   |                  |
| 7         | S7                   | S3                     | S10              | S6               |                   |                  |
| 8         | S10                  | S5                     | S2               | S7               |                   |                  |
| 9         | S6                   |                         | S1               |                   |                   |                  |

Discovery of knowledge by analysis process of acquired results could be determined using the estimation accuracy on the same information by application of tree decision for all three departments, in the case of Relieff PrecR(k) classifier application compared to the aggregation method PrecA(k) has expressed better results. The number of attributes k was less, and estimation accuracy Prec(k) is better:

- Criminology Department: estimation accuracy is 84.5 with only 4 relevant attributes -PrecR(4) = 84.5% comparing to the accuracy of 82.5 with the set of 9 attributes-PrecA(9) = 82.5% ;
- Police Department: estimation accuracy is 93.8 with the set of 5 attributes-PrecR(5) = 93.8% comparing to the accuracy of 83.0 with 8 relevant attributes-PrecA(8) = 83.0%;
- In the case of the Department for Safety, the situation is different. Higher precision of 88.3 was achieved by the set of 3 attributes-PrecR(3) = 88.3% extracted on the basis of their coefficients acquired by Relieff classifier than the set of 9 relevant attributes extracted by aggregation of 83.0-PrecA(9) = 83.0%;
Comparing sets of extracted attributes, we can observe that both methods as the same attributes give: Police Department-8, Department for Safety-1 and Criminology Department-7.

However, if we take a look at the smallest difference of ε achieved in the mid-value of estimation accuracy by applying attributes ranking by aggregation PrecA(k) and application of ReliefF classifier PrecR(k), we have accomplished the following aberrance: Department for Safety’s value is $\varepsilon = 0.022$ (for the set of 4 attributes), Criminology Department’s value is $\varepsilon = 0.001$ (for set of 8 attributes) and Police Department’s value is $\varepsilon = 0.000$ (for set of 9 attributes). By comparing both sets of attributes with all three departments, we could extract the set of the same attributes that represent relevant modules when selecting the Department at the CPA. In Table 15 (marked with gray), it is clearly visible that S8-Constitutional law, S2-English language (excluding the case of application of ReliefF classifier to Safety department) and S5-Introduction to the law (excluding the case of application of aggregation to Safety department) are important for all three departments, that the modules S3-Criminal law general part, S7-Police equipment and S10-IT are important for the Police Department and Criminology Department. For the Safety Department and Police Department (excluding the case of application of aggregation to the Safety Department), module S1-Fundamentals of economics is also important.

5. Conclusions

The main hypothesis of the authors was to show the possibility of the development of one aggregated model having better characteristics than integrated techniques when they are used independently. The techniques which were considered in this paper solved the problem of determination and prediction of the importance of attributes that describes one multidimensional and multicriteria process. The evaluation was carried out on the case study to the determination of students’ future behavior regarding the choice of one of the three offered modules at the Criminal and Police University, Belgrade. The method of feature selection was used to reduce the number of attributes, thus increasing the accuracy of the model. As traditional techniques, the authors demonstrated the usage of the method of regression analysis combined with the AHP mathematical method. The application of ReliefF classificatory ranking method of machine learning was used together with decision tree classifiers J48 for aggregation purposes. The evaluation results have shown the advantage and supremacy of the proposed model. The authors have claimed that the proposed model has no significant limitations, and we will consider the inclusion of n-modular redundancy into it in our future work related to this topic.

It is important to mention that the obtained analysis of the personal contribution of the students via achieved results on the first-year subject modules can help in the process of achieving relevant attributes, which are, in this case, related to the determination of the robustness of the first-year modules that could be prearranged by the experts at the Ministry of Interior, The Republic of Serbia as well as adequate support in determination of toughness that is given to the first-year modules through the subject accreditation process with a corresponding number of European Credit Transfer and Accumulation System (ECTS) credits.

Author Contributions: Conceptualization, Investigation, Validation: M.R. and A.A.; Project administration: R.R. and V.S.; Writing—review & editing, Formal analysis, Software: M.Č.; Supervision, Methodology, Writing—original draft: D.R. All authors have read and agreed to the uploaded version of the manuscript. All authors agreed in advance version of the manuscript which eventual will be published. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.
Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank the Nis Science and Technology Park for their support in the publishing of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Hart, A. Machine Induction as a Form of Knowledge Acquisition in Knowledge Engineering; Machine Learning: Principles and techniques; Chapman and Hall: London, UK, 1989.
2. Haiyang, Z. A Short Introduction to Data Mining and Its Applications. In Proceedings of the Management and Service Science (MSS’11), International Conference IEEE, Wuhan, China, 12–14 August 2011; pp. 1–4.
3. Saaty, T.L. Multicriteria Decision Making: The Analytic Hierarchy Process; McGraw-Hill: New York, NY, USA, 1980.
4. Vaidya, O.S.; Kumar, S. Analytic hierarchy process: An overview of applications. Eur. J. Oper. Res. 2006, 169, 1–29.
5. Harrell, F.E. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis; Springer: New York, NY, USA, 2001.
6. Montgomery, D.C.; Peck, E.; Vining, G. Introduction to Linear Regression Analysis, 6th ed.; Wiley: New York, NY, USA, 2021.
7. WEKA Software; The University of Waikato: Hamilton, New Zealand, 2015. Available online: http://www.cs.waikato.ac.nz/ml/weka (accessed on 30 May 2022).
8. Green, B.; Johnson, C.; McCarthy, C. Predicting Academic Success in the First Year of Chiropractic College. J. Manip. Physiol. Ther. 2003, 6, 40–46.
9. Rothstein, C. College performance predictions and the SAT. J. Econom. 2004, 121, 297–317.
10. Montmarquette, C.; Mahseredjian, S.; Houle, R. The determinants of university dropouts: A bivariate probability model with sample selection. Econ. Educ. Res. Rev. 2001, 20, 475–484.
11. Huang, S.; Fang, N. Regression models of predicting student academic performance in an engineering dynamics course. In ASE Annual Conference and Exposition, Conference Proceedings; American Society for Engineering Education (ASEE): Louisville, KY, USA, 2010; ISBN:9781617820205; pp. 151026.1–151026.17. https://doi.org/10.18260/1-2-15759.
12. Ayan, R.; Noel, M.; García, M. Prediction of University Students’ Academic Achievement by Linear and Logistic Models. Span. J. Ophthalmol. 2008, 11, 275–288. https://doi.org/10.1017/S1138741600004315.
13. Saaty, L.T. Analytical Planning: The Organization of Systems; Pergamon Press: Oxford, UK, 1985.
14. Tadisina, S.K.; Troutt, M.D.; Bhavis, V. Selecting a Doctoral Programme Using the Analytic Hierarchy Process—The Importance of Perspective. J. Oper. Res. Soc. 1991, 42, 631–638.
15. Delibašić, B.; Suknović, M.; Stanačev, D. Menadžment Znanja Pri Izboru Odgovarajućeg Smera na Studijama; INFO-M, 2005; Volume 4, br. 14, str. 23–27. Available online: https://scindeks.ceon.rs/article.aspx?artid=1451-43970514023D (accessed on 30 May 2022)
16. Savić, G.; Makajić-Nikolić, D.; Randelović, D.; Randelović, M. Study program selection by aggregated DEA-AHP measure. Metal. Int. 2013, 1, 169–174.
17. Strayhorn, T.L. An examination of the impact of first-year seminars on correlates of college student retention. J. First Year Exp. Stud. Transm. 2009, 21, 9–27.
18. Kovačić, Z.I. Predicting student success by mining enrolment data. Res. High. Educ. J. 2012, 15, 1–20.
19. Rajput, H.; Milani, A.; Labun, A. Including time dependency and ANOVA in decision making using the revised fuzzy AHP: A case study on wafer fabrication process selection. Appl. Soft Comput. 2011, 11, 5099–5109.
20. Sandra, L.; Lumbangao, F.; Matsuo, T. Machine Learning Algorithm to Predict Student’s Performance: A Systematic Literature Review. TEM J. 2021, 10, 1919–1927.
21. Yakubu, M.N.; Abubakar, A.M. Applying machine learning approach to predict students’ performance in higher educational institutions. Kybernetes 2021, 51, 916–934.
22. Lynn, N.D.; Emanuel, A.W.R. Using Data Mining Techniques to Predict Students’ Performance. a Review. IOP Conference Series: Materials Science and Engineering. In Proceedings of the 6th International Conference on Industrial, Mechanical, Electrical and Chemical Engineering—ICIMECE 2020, Solo, Indonesia, 20 October 2020; Volume 1096.
23. Siddique, A.; Jan, A.; Majeeed, F.; Qahmash, A.I.; Quadri, N.N.; Wahab, M.O.A. Predicting Academic Performance Using an Efficient Model Based on Fusion of Classifiers. Appl. Sci. 2021, 11, 11845. https://doi.org/10.3390/app112411845.
24. Buenaño-Fernández, D.; Gil, D.; Luján-Mora, S. Application of Machine Learning in Predicting Performance for Computer Engineering Students: A Case Study. Sustainability 2019, 11, 2833. https://doi.org/10.3390/su11102833.
25. Hwang, G.J.; Xie, H.; Wah, B.; Gasevic, D. Vision, challenges, roles and research issues of Artificial Intelligence in Education. Computers and Education: Artif. Intell. 2020, 1, 10001. https://doi.org/10.1016/j.cai.2020.10001.
26. Baars, G.J.A.; Stijnen, T.; Splinter, T.A.W. A Model to Predict Student Failure in the First Year of the Undergraduate Medical Curriculum. Health Prof. Educ. 2017, 3, 5–14. https://doi.org/10.1016/j.hpe.2017.01.001.
61. Amrieh, E.A.; Hamtini, T.; Aljarah, I. Mining Educational Data to Predict Student’s academic Performance using Ensemble Methods. *Int. J. Database Theory Appl.* **2016**, *9*, 119–136.

62. VCO Software (2015), Faculty of Organizational Sciences. University of Belgrade. Available online: [http://www.odlucivanje.fon.rs/images/stories/download/VKO.zip](http://www.odlucivanje.fon.rs/images/stories/download/VKO.zip) (accessed on 30 May 2022).

63. SPSS Inc. *Released 2008. SPSS Statistics for Windows, Version 17.0*; SPSS Inc.: Chicago, IL, USA, 2008.

64. Wellman, B. *Doing It Ourselves: The SPSS Manual as Sociology’s Most Influential Recent Book; Required Reading: Sociology’s Most Influential Books*; University of Massachusetts Press: Amherst, MA, USA, 1998; pp. 71–78.

65. Vatcheva, K.P.; Lee, M.; McCormick, J.B.; Rahbar, M.H. Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies. *Epidemiology (Sunnyvale)* **2016**, *6*, 227.

66. Shrestha, N. Detecting Multicollinearity in Regression Analysis. *Am. J. Appl. Math. Stat.* **2020**, *8*, 39–42.

67. Santafe, G.; Inza, I.; Lozano, J.A. Dealing with the evaluation of supervised classification algorithms. *Artif. Intell. Rev.* **2015**, *44*, 467–508.

68. Dimić, G.; Prokin, D.; Kuk, K.; Micalović, M. Primena Decision Trees i Naive Bayes klasifikatora na skup podataka izdvojen iz Moodle kursa. In Proceedings of the Conference INFOTEH, Jahorina, Bosnia and Herzegovina, 21–23 March 2012; Volume 11, pp. 877–882.

69. Yang, S.; Berdine, G.. The receiver operating characteristic (ROC) curve. *Southwest Respiratory Critical Care Chronicles* **2017**, *5*, 34.

70. Vuk, M.; Curk, T. ROC curve, lift chart and calibration plot. *Metodoloski Zv.* **2006**, *3*, 89–108.