Received: 8 June 2018 | Accepted: 20 October 2018
DOI: 10.1002/cae.22092

RESEARCH ARTICLE

WILEY

Using the Belbin method and models for predicting the academic performance of engineering students

Luis Gutiérrez | **Víctor Flores** | **Brian Keith** | **Aldo Quelopana**

Department of Computing & Systems Engineering, Universidad Católica del Norte, Antofagasta, Chile

Correspondence
Víctor Flores, Department of Computing & Systems Engineering, Universidad Católica del Norte, 0610 Angamos Av., Antofagasta, Chile.
Email: vflores@ucn.cl

Abstract

This paper describes the process of generating a predictive model of students' academic performance in different engineering subjects at Universidad Católica del Norte (UCN). It aims to analyze the importance of variables influencing the final average grade of the UCN students in projects related to different subjects, focusing on the dimensions resulting from the Belbin test. The main objective of this work is to provide evidence of the real impact of the Belbin test outcomes on the final performance of a team of students, using as a metric of variable importance the one provided by a Random Forest model, supplied by the *scikit-learn* library. As a result, the final classifier presented an accuracy of 80%, and one of the most influential variables according to this model was Covered Roles 2, which represents the number of roles covered in each group. Future research lines are proposed to validate these outcomes, mostly concerned with the acquisition of more data across several future semesters.

KEYWORDS

active methodologies, belbin questionnaire, random forest, variable importance

1 | INTRODUCTION

Some challenges have been recently considered in college education, such as generating significant changes in the way teaching is done or training students to behave properly in a sustainable professional world. These challenges lie within the declaration of the European Higher Education Space, as a student-centered paradigm.[1]

To face these challenges appropriately, it is necessary to develop competencies and train academics, but it is also important to develop the students' cognitive and non-cognitive skills. According to recent studies, students need these skills to meet their academic expectations [17]. Among the required skills are soft ones, including work ethics, positive attitude, communication, social skills, learning motivation, and teamwork skills [7,13].

Soft skills complement professional knowledge and allow, as a whole, students and future workers to perform better in their work environment [13]. Studies have been conducted on how emotional intelligence and certain personality features are positively related to students' academic success, at least in an online environment [5].

In engineering majors and particularly project development, it is increasingly more important for an academic to identify students' potentialities to develop soft skills such as the ones mentioned above. In activities such as project development, significant achievements in both technical skills and soft skills have been reported [9].

In the new teaching and learning scheme, the student has somehow changed from a passive element that receives contents into the center of the teaching and learning process, thus requiring that students put into practice skills associated with their profile, which they must have developed during pre-college training.

On the other hand, a professor is a good professional who masters his discipline, but with little systematic training to allow him teaching and identifying students' skills so that these can work on a success-oriented approach, according to the recent student-centered paradigm.

The literature contains abundant papers to study/propose the features and desirable profile for an academic so that he can achieve learning outcomes in a better way. But papers focusing on efforts to identify a student's profile characteristics in order to improve their academic performance are not so common. This paper shows the results of an effort to identify students' characteristics so that they can appropriately face teamwork tasks focusing on engineering projects.

Many times, work groups are formed by students' affinity and, other times, they are organized by professors. With the purpose of taking advantage of students' potentialities, this study uses the Belbin method, which provides a student profile facilitating the assignation of tasks in the work groups. The results of the application of the Belbin method and predictive model are described in detail in the document, along with the results obtained and possible future research lines.

2 | CONTEXT

UCN has declared its Institutional Educational Project[2] within the framework of competency-based teaching from a social constructivist paradigm which involves assuming the new student-centered learning paradigm [9]. In this context, teaching in UCN engineering majors has benefited from the application of new teaching techniques such as teamwork.

This experience has been developed at UCN Faculty of Engineering and Geological Sciences, with two objectives: first, identify students' teamwork skills through the Belbin method and observe students' performance outcomes through the grade obtained after the group work is finished; and second, after grading the team, create a predictive model based on artificial intelligence techniques to predict academic performance for new cases.

The Belbin model and the predictive models, particularly the Random Forest Model, are described below, along with the directions for both techniques to be used for teamwork in college teaching.

2.1 | The Belbin method for teamwork

The Belbin methodology was proposed by Meredith Belbin after a 6-year experimental study. It consists of a series of psychometric tests and a questionnaire to evaluate critical thinking. Its aim is to facilitate cooperation and teamwork through the assignation of roles to each team member [3]. This is done under the premise that teams formed heterogeneously in terms of roles are more efficient than those who are not, thus proposing that, more than intellectuality, the balance of roles among members is the most important for team success.

The Belbin method is basically a technique for identifying team members' profiles. It is based on the idea of group work roles. These roles deal with the distribution of tasks and responsibilities within the work team [1]. The general characteristics of Team Roles can be briefly described as follows [3,4]:

- **Plant:** creative, imaginative, unorthodox. Solves difficult problems.
- **Resource Investigator:** extrovert, enthusiastic, communicative. Looks for new opportunities.
- **Coordinator:** mature, self-confident. A good director of work teams. Promotes decision-making.
- **Shaper:** dynamic, thrives on pressure.
- **Monitor-Evaluator:** sober, discerning, strategic. Identifies options in task development.
- **Team-worker:** cooperative, perceptive, and diplomatic. Listens and averts frictions.
- **Implementer:** disciplined, reliable, and efficient. Turns ideas into practical actions.
- **Completer-Finisher:** painstaking, conscientious, and anxious. Good for searching out errors and omissions to correct them.
- **Specialist:** Interested in accomplishing duties and intends to do one task at a time. Provides specific knowledge.

The Team Roles described above are used to improve the teams' interactions. Teamwork is a technique used in entrepreneurial and educational environments for collective purposes with individual contributions [23]. In a teaching context, teamwork is a highly powerful technique, particularly in engineering majors, apart from being a competency quite valued by employers. On a college basis, teamwork is a suitable platform for developing values and skills to "behave" at work with other people [10].

This paper does not consider the role of a specialist in the questionnaire because all students in a team could be regarded as specialists. This is because all students may be considered specialists in their major's professional area. So, only eight of the nine roles proposed by Belbin are considered, the skills of each role and how to determine them are described in further detail in the work of [4].

2.2 | Predictive models

Predictive models use statistics for predicting outcomes [12]. In general, the event to predict is in the future; however, predictive modeling can be applied to an unknown event, regardless of the time when it occurred.

The use of predictive models and data analysis in higher education is still a relatively new area in both the academic environment and its practical application [27]. Professors could eventually use new data sources available for directing subject redesign and as evidence for implementing new types of assessment and communication channels to strengthen teacher-student relationships [2].

By making use of students' data, higher education institutions can build statistical and machine learning models to predict students' outcomes. For example, a case study in a community college used data analysis and predictive modeling to identify students in risk, based on a series of key variables [26].

2.3 | Random Forest model

Random Forest (RF) models are a type of predictive model based on decision trees, in particular, RF is a general technique of random decision trees. This method combines the idea of *bagging* with a random selection of characteristics, with the intention of building decision trees with controlled variance [14,15]. It is an *ensemble method* for classification and regression tasks, which operates through the construction of multiple decision trees during training. The method consists of generating several trees using random subsets and then combining the results of these independent trees to obtain the final result. In the case of classification, the determined class corresponds to the mode of the classes provided by each tree. In the case of regression, it corresponds to the average prediction of individual trees. Random decision trees correct the decision tree tendency to overfit their training set [11].

The RF model was implemented in the *scikit-learn* library [21], initially in its regression version. This model is convenient for solving the problem described in this paper because it immediately provides a way to measure the importance of each variable in its implementation. In fact, one of the main benefits of RFs as models is that they can be used to determine the importance of variables in a regression or classification problem in a natural way [6]. The importance of each variable is calculated with a metric, based on the impurity decrease in each node used for during the data partition process of the decision tree.

One of the reasons for using RFs over other predictive models is that RFs and ensemble methods in general, are usually the classifiers/regressors rendering one of the best *out-of-the-box* results [8]. There are various studies that have used classical Decision Trees to predict student performance, because of its simplicity and comprehensibility to uncover small or large data structure and predict the value. Given that Random Forests are fundamentally an Ensemble of Decision Trees with usually higher predictive power, it seems natural to use these models [25]. Also, RFs have been previously applied in the literature to determine the importance of the factors associated with students' retention in science and engineering majors [18]. Therefore, there is a precedent for its application in the educational area.

3 | METHODOLOGY AND EXPERIENCE

The methodology used consisted of seven stages, there are described below. The sample consisted of 245 students divided into 49 teams of different sizes (2–5), from 9 engineering subjects selected for this experience, particularly: Capstone Project, Information Systems II, Automata Theory and Formal Languages, Basic Computing, Introduction to Programming, Programming Workshop, Project Evaluation, and Software Engineering II.

The teams belonged to the majors of Industrial Engineering (ICI), Computing and Informatics Engineering (ICCI), Metallurgy Engineering (IEM), Chemical Process Engineering (IEPQ), Geology, Metallurgical Engineering (ICM), Civil Common Curriculum Engineering (ICPC), and Computing and Informatics Engineering (IECI, note that this is different from the previous one and is more technically oriented). Team frequency per major is shown in Table 1. Each stage of the methodology is shown below:

- **Administering Belbin questionnaire:** In this stage, each student in a subject completed the questionnaire. Also, data collected was kept in a digital format.
- **Analyzing results:** The results of each questionnaire were analyzed by a UCN teaching team. In particular, the analysis consisted in identifying the main role of each student and establishing what roles showed the greatest frequency in the different subjects.
- **Assigning roles to teams:** This stage consisted of assigning responsibilities to teams, according to the most frequent role shown in the results. It was based on the identified roles by students in the previous task in order to

TABLE 1 Frequency of majors

Major	Frequency
ICI	29
ICCI	4
IEM	4
IEPC	1
Geology	3
ICM	2
ICPC	3
IECI	3

take advantage of the potentialities of each student in their teams.

- **Collecting and classifying results:** Once final grades were obtained for each subject, this stage consisted in classifying students' performance using the following labels: insufficient (for grades lower than 4, the lowest passing grade in Chile), sufficient (for grades between 4 and 5.0), regular (for grades between 5.1 and 5.8), good (for grades between 5.9 and 6.6), and excellent (for grades between 6.7 and 7, the highest passing grade in Chile). These values were obtained by looking for a uniform data distribution in the different teams. In addition, no grades in the insufficient range were reported in the subjects studied.
- **Generating a predictive model:** In this stage, data collected and generated were used for creating a predictive model for academic performance, according to the roles assigned to teams.
- **Comparing results:** This stage consisted in comparing the results obtained from subjects, using the predictions of the predictive model.

In the stage for collecting and classifying results, the following variables were considered for each team: name of the team, subject, major, average age of the team, team size, number of members. Variables whose meaning is not immediate are listed in Table 2.

By considering the variables and their meanings above, those variables that could better contribute to the model were selected. The final selection of variables consisted of Subject, Major, Term (Semester), Age, Team Size, Average Teamwork Hours, Average Coordination Hours, Average Individual Hours, Team Communication, Covered Roles 2, and Covered Roles 3.

In order to provide an adequate mathematical representation of "Covered Roles n" (and in particular Covered Roles 2 and 3) set theory was used. Let S be the set of possible roles. Each student x defines a total order R_x on the set S. We then define the set of the top n roles for a student x recursively as

$$T_x(n) = \begin{cases} \max\limits_{R_x}(S), & if n = 1 \\ \max\limits_{R_x}(S - T_x(n-1)), & if n \geq 1 \end{cases}$$

Where $\max\limits_{R_x}(A)$ refers to finding the maximum element in the set A with respect to the total order R_x. Using this definition, it is possible to define "Covered Roles n" unambiguously for a set of students G (representing a team of students) as the cardinality of the set

$$C_n = \bigcup_{x \in G} T_x(n).$$

For illustration purposes, consider a group of three students A, B, C, and only five roles numbered one to five. The ranking of roles of each student is presented in Table 3.

If we want to calculate Covered Roles 2 we would obtain the following sets:

TABLE 2 Description of the variables in the collected data

Variable	Meaning
Number of the team	An auto-incrementing number that refers to the team identifier.
Date	The date of end of subject.
Term (semester)	The term in which the subject is taught, relative to the major's program of study.
Average teamwork hours	The average number of hours of teamwork. This average, the same as the following ones, is calculated according to the hours each member reports as work time.
Average coordination hours	The average number of hours the team dedicated to coordination
Average individual hours	The average number of hours the team dedicated to individual work (i.e., the time each individual dedicated to work, as a team average)
Team duration	The duration of the team in days.
Complexity	The project complexity (calculated as the sum of the average team hours, average coordination hours, and average individual hours, multiplied by the team duration, divided by 7).
Team communication	The team communication is an estimated value given by each member, representing the quality of communication throughout the project, taking values from 0 to 1 or 0 to 100, where 0 is bad communication and 100 is excellent communication. This value is then averaged to obtain the final team communication.
Covered roles 2	A number indicating all the Belbin's roles covered in a team, considering only the first two roles of each member.
Covered roles 3	The same as Covered Roles 2, but with 3 roles of each member.
Coefficient 2	A coefficient calculated as Covered Roles 2, divided by team size.
Coefficient 3	A coefficient calculated as Covered Roles 3, divided by team size.

TABLE 3 Example group for the calculation of Covered Roles *n*

	Role				
Student	**Role 1**	**Role 2**	**Role 3**	**Role 4**	**Role 5**
A	3	1	2	4	5
B	5	2	3	1	4
C	3	4	1	2	5

The numbers in the cells represent the ranking of each role with lower values representing a higher ranking.

$$T_A(2) = \{2,\ 3\}$$

$$T_B(2) = \{4,\ 2\}$$

$$T_C(2) = \{3,\ 4\}$$

Thus, the union of this set $C_2 = \{2{,}3{,}4\}$, and thus Covered Roles 2 would take the value 3. On the other hand, if we wanted to calculate Covered Roles 3 we would obtain the sets:

$$T_A(3) = \{2,\ 3,\ 1\}$$

$$T_B(3) = \{4,\ 2,\ 3\}$$

$$T_C(3) = \{3,\ 4,\ 1\}$$

Thus, the union of this set $C_3 = \{1{,}2{,}3{,}4\}$, and thus Covered Roles 2 would take the value 4.

While Covered Roles 3 was originally included in the analysis of the Belbin questionnaire results, it was later excluded for the purposes of the predictive model. This was done because a connection was expected between Covered Roles 2 and the final grade, thus Covered Roles 3 appeared to be redundant. In addition, other variables that were closely related to another were also excluded, in the sense that they can be calculated as a function of the other variables available. Finally, the initial hypothesis is that roles covered according to Belbin (as represented by Covered Roles 2) deeply influence the final grade of group projects [3]. Thus, variable Covered Roles 3 was excluded from the final model.

4 | EVALUATION

Initially, the problem was considered as a regression in order to obtain a numeric prediction of the grades from the data of each team. From this analysis, a model based on RF was obtained and the importance of associated variables was studied. Results and details are shown below.

We first perform a parameter tuning phase and then evaluate our proposed model using a 40-fold cross-validation 10 times and then averaging the final results. After the tuning process our choices of parameters for our random forest classifier, as defined in the *scikit-learn* library Pedregosa et al. (2011), were given by a number of Estimators (trees) of 40. The other parameters use their default value and behavior. With this approach, we obtain a negative R^2 score (in most cases less than -1), which indicates that the model is performing worse than the simplest possible estimation for regression: a constant line representing the mean of the output variable. Since this kind of model would have an R^2 score of zero.

Even though the results of cross-validation for regression have not been good, for completeness. We proceed to use our defined optimal parameterization on a general model that takes 80% of the samples for training and 20% of the samples for testing (i.e., a hold-out validation) to evaluate a final model and to obtain our final variable importance measures. Running this model 10 times and averaging the results gives an average R^2 score of 0.4812, which is many times better than the negative scores (which were usually less than -1).

We now train a full model using all the data available to calculate the variable importance and to give example predictions. Table 5 shows the new variable importance, after discretizing. Table 4 shows the predictions after discretizing.

TABLE 4 Comparison between predicted and real values, using the RF regression model (results for the best regression model with an R^2 of 0.7873)

Subject	Major	Predicted	Real
Basic computing	Metallurgical Engineering	5.56	5.2
Project evaluation	Industrial Engineering	5.97	5.8
Basic computing	Geology	5.34	4.9
Introduction to programming	Industrial Engineering	6.49	7.0
Introduction to programming	Metallurgical Engineering	6.68	7.0
Capstone project	Industrial Engineering	5.24	5.2
Introduction to programming	Industrial Engineering	6.68	6.8
Introduction to programming	Civil Engineering Common Core	6.70	6.6
Project evaluation	Industrial Engineering	6.04	6.5
Capstone project	Industrial Engineering	5.57	5.8

TABLE 5 Model variables and their importance, from the greatest to the smallest for the best model with regression random forests ($R^2 = 0.7873$)

Variable	Importance
Average Individual Hours	0.335139
Age	0.220918
Subject	0.176173
Team Communication	0.064718
Covered Roles 2	0.062782
Team Size	0.058621
Major	0.035544
Term (semester)	0.028080
Average Team Hours	0.009883
Average Coordination Hours	0.008141

As to the analysis of variable importance, the *scikit-learn* library, with which this study was implemented, uses the *mean decrease impurity* metric [21] to determine the variable importance of the problem. This is defined as the total impurity decrease of a node, weighed with an estimation of the probability for reaching this node, considering all the RF trees [6]. Results shown in Table 5 are obtained by adjusting the model.

The variable with the greatest importance is Average Individual Hours. This may be explained by considering the relevance of individual efforts so that the teams can reach their best performance. In this context, other variables of importance are Age, giving evidence of a possible correlation of performance to emotional aspects such as group maturity; and Subject, which may be related to the inherent difficulty of the associated project. The importance of the other variables is lower than 20%. Table 5 shows that Covered Roles 2 ranks in the fifth place, separated from the first one by a 0.2732 difference. These results can be visualized in Figure 1.

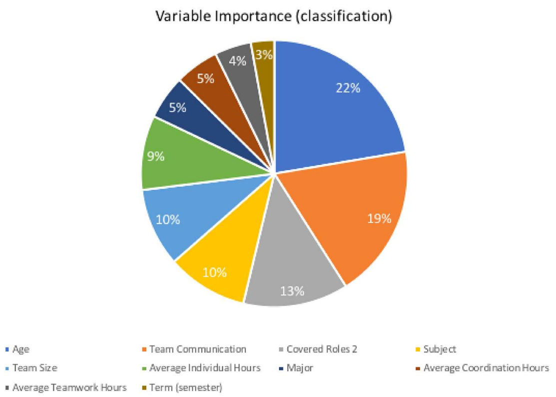

FIGURE 1 Pie chart showing the variables and their importance for the best model with regression random forests

Table 4 shows that, when predicting, the model does not reveal a good performance. This is formally shown through the performance metric, R^2, which represents the data variance ratio explained by the model, where the best possible value for R^2 is 1.0 [2]. In the context of this regression problem, R^2refers to the variance in grades that can be explained using this model. In particular, for this model and using the values of the hold-out case, at best, the value of R^2 is 0.7873. However, given the results during the cross-validation phase, which returned negative values of R^2, this suggests that a regression model may not be a good approach to this problem.

For a better performance, the problem was transformed from a regression one to a classification one through grade discretization, as described above. Then, each grade was labeled according to its range. The ranges and labels are shown in Table 6.

We first perform a parameter tuning phase and then evaluate our proposed model using a 40-fold cross-validation 10 times and then averaging the final results. After the tuning process our choices of parameters for our random forest classifier, as defined in the *scikit-learn* library Pedregosa et al. (2011), were:

- Number of Estimators (trees): 100
- Criterion: Entropy
- Max_Depth: 10
- Bootstrap: False

The other parameters use their default value and behavior. With this approach we obtain the following results:

- F1 scores average: 55.67%
- Accuracy average: 52.83%

Having obtained good results using a cross-validation scheme, we proceed to use our defined optimal parameterization on a general model that takes 80% of the samples for training and 20% of the samples for testing (i.e., a hold-out validation) to evaluate a final model and to obtain our final variable importance measures. Running this model 10 times and averaging the results gives an average F1 score of 59.10%.

TABLE 6 Grade range and their corresponding labels

Range	Label	Frequency
$1 \leq$ Grade <4	Insufficient	0
$4 \leq$ Grade ≤ 5	Sufficient	11
5 Grade ≤ 5.8	Regular	13
5.8 Grade ≤ 6.6	Good	13
6.6 Grade ≤ 7	Excellent	12

We now train a full model using all the data available to calculate the variable importance and to give example predictions. Table 7 shows the new variable importance, after discretizing. Table 8 shows the predictions after discretizing. Figure 2 shows a visualization of variable importance for the classification model.

Unlike Table 5, after discretizing the problem, variable importance shows more uniform values. The variable with the greatest importance in the discretized problem is the Age of the students, with Team Communication in second place. Furthermore, the Covered Roles 2 variable is the third most important variable, thus suggesting a significant relationship between the coverage of the roles defined by the Belbin method and the category of the final grade of the project. These results can be visualized in Figure 2.

The real results of each subject (discretized) are compared with the results obtained for the model in Table 8. The details of this comparison are shown in Table 8, which was built in the same way as Table 4.

The predictions obtained for each team could be used as an indicator of the teams that could potentially need greater support during the development of their projects in each subject. Particularly, if in the initial stages of a project the data for a specific team is fed and the prediction is obtained in the Sufficient or Regular range, some kind of a remedial and supportive measure could be applied to improve their performance so that they can approach the Good or Excellent range. In this context, the model can be used as a support for the teaching and learning process of a subject.

For evaluation purposes, the standard metrics of accuracy and the F_1score are reported. Also, precision and recall (the two constituents of the F_1score) are explained. A brief description of each metric and its interpretation is given here, based on the work by [24]. Accuracy refers to the ratio between the correctly predicted cases and the total number of cases, in this problem it refers to the correctly predicted grade ranges. On the other hand, precision refers to the fraction of relevant instances among the retrieved instances. In this case, it would refer to the likelihood of a grade being of a certain class C given that it was predicted to be of class C. On the other hand, recall refers to the the fraction of relevant instances retrieved over the total number of relevant instances in the problem. For this problem, recall would refer to the likelihood of a member of class C to be retrieved over all the members of class C. The F_1 score is the harmonic mean of precision and recall, and thus it can be used as a balanced measure of the performance of a classifier which combines both the information of precision and recall.

Apart from obtaining evidence of a possible relationship between the covered roles defined by Belbin and the category of the final grade of the project. Of all the models trained, the best one, we can conclude that the classifier performance is competitive in itself, with a value of F_1 of 80.00% and an accuracy of 80.00% too, which places it highly over the baseline of a random classification. Table 9 shows the confusion matrix with classification results for the best model. While these are the results for the best classifier reported, the average results shown previously also show that the model is highly competitive.

TABLE 7 Variable importance after discretizing obtained with the best model for classification with random forests (accuracy of 80%)

Variable	Importance
Age	0.22141
Team Communication	0.18456
Covered Roles 2	0.1264
Subject	0.09701
Team Size	0.09405
Average Individual Hours	0.08944
Major	0.05279
Average Coordination Hours	0.05276
Average Teamwork Hours	0.04385
Term (semester)	0.03773

TABLE 8 Predicted values for the discretization obtained with the best model for classification with random forests (accuracy of 80%)

Subject	Major	Predicted	Real
Basic Computing	Metallurgical Engineering	Regular	Regular
Programming Workshop	Computer Engineering	Sufficient	Sufficient
Introduction to Programming	Civil Engineering Common Core	Good	Excellent
Introduction to Programming	Metallurgical Engineering	Excellent	Excellent
Introduction to Programming	Civil Engineering Common Core	Good	Good
Basic Computing	Metallurgical Engineering	Regular	Regular
Programming Workshop	Computer Engineering	Sufficient	Sufficient
Automata Theory	Computer Engineering	Regular	Sufficient
Introduction to Programming	Metallurgical Engineering	Excellent	Excellent
Project Evaluation	Industrial Engineering	Good	Good

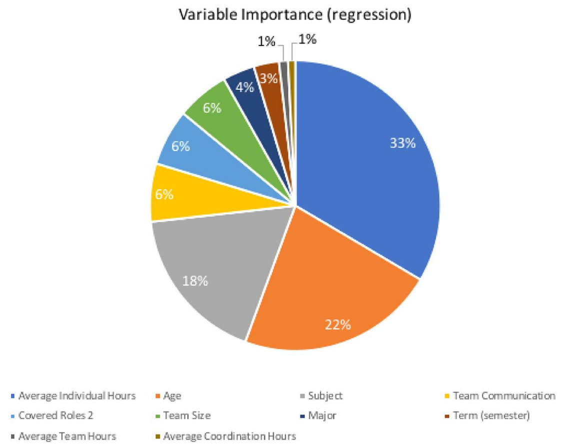

FIGURE 2 Pie chart showing the variable importance after discretizing obtained with the best model for classification with random forests

5 | DISCUSSION

The differences in the results of variable importance between the regression model and the classification model could be due to the worse results of the regression model. Especially since when attempting to apply a cross-validation scheme to the regression models no good models could be constructed according to the R^2 criteria. In contrast, the classification scheme provides decent results considering it is a multiclass problem. In this context, it would be expected that the results from the classification problem to be more reliable. Thus, the difference between these two models could be accounted for because of their different predictive capabilities.

It should also be noted that in the context of grade prediction, an exact grade is not necessarily needed, rather an estimation of the range of the student performance is usually sufficient. Thus, the classification approach seems more natural to this problem. This intuition is reinforced by the relatively better results when using the random forests for classification rather than regression.

The aim of the Belbin method is to improve the performance on teamwork through an adequate assignation of roles. This method has been used successfully in several academic and industrial projects, as shown in [3]. In this work, we observed an apparent relationship between Team Communication and the assignation of roles (Covered Roles 2) using the Belbin method. This relationship must be corroborated with further studies.

Based on our results (shown in Table 7), it could be possible to think of a relationship between the variables Age, Team Communication and Covered Roles 2 that benefits the team performance. In a future work, some Artificial Intelligence techniques (such as Bayesian Networks) could be used, in order to identify conditional relationships among the different variables considered in this work. As an initial hypothesis, Age and Covered Roles have a positive influence on Team Communication. This hypothesis could be validated on a future work using more data.

In general, there are several works that report predictive models for students' performance, some of these present particular models to improve student performance. For example, decision trees are used on the works of [22], [19], [20], and [16] with accuracies between 65 and 91%. The accuracy of our model falls in this range, thus it offers a competitive performance, although it should be noted that the attributes used in each work are different.

According to [25], there are several studies using data mining techniques to predict student performance, most of these studies use variables such as the Grade Point Average and other internal measures of student performance. While other studies also apply demographic information, such as family background and gender. In contrast, this study focuses on the usage of the Belbin test results to predict the performance of a whole group.

TABLE 9 Confusion matrix with classification results

	Sufficient	Regular	Good	Excellent
Sufficient	2	1	0	0
Regular	0	2	0	0
Good	0	0	2	0
Excellent	0	0	1	2

6 | CONCLUSIONS AND FUTURE RESEARCH LINES

The main purpose of this paper is the development of a model to predict the student team's performance. This was done by collecting general data associated with work teams and data provided by Belbin questionnaire. These data allowed building a predictive model using a Random Forest model for regression and classification, where the classification model obtained a better performance. This predictive model allows determining teams that have a higher risk (in terms of getting a lower grade) in a subject. In this context, the teams characterized in the Sufficient or Regular range could give additional support so that they can reach a better performance.

Our results discussed in the sections above could be verified with additional results in future works. In particular, relationships between variables could be studied in future works using Artificial Intelligence techniques (such as Bayesian Networks). Another outcome places Covered Roles, according to Belbin questionnaire, as an aspect with a great impact on team performance, classifying performance into academic categories.

Other variables also regarded as relevant are team age or the subject in which final projects were included. In addition, the performance of classification using these variables is 80% accuracy as shown in Table 7. According to our analysis, the most influential variables in this model were Age, Team Communication and Covered Roles 2 it is closely related to the use of the Belbin method for assigning the roles of the teams.

Furthermore, the study of the influence of team roles proposed by Belbin is still open for more study. In particular, capturing more data than the 49 original records for future research would be the first approach. This may allow analyzing the influence of these factors and how they affect evaluation more appropriately, apart from testing if the validity of the classification model based on Random Forests is adequate or not for a work domain. For future works, we will make more surveys over successive semesters and their results will be analyzed, using the model generated in this proposal, in order to validate the model performance and draw new conclusions.

As a complementary future line, data from the same subjects are likely to be captured (Table 4), but with new teams of students, during the two following academic years. Another future line would be capturing data using Belbin questionnaire to analyze and apply the model in subjects from majors different than those considered in this study in order to validate the model generated and described in this paper again.

Finally, while more powerful models, such as deep neural networks, could be applied. Given that there is not much data to work with, using such complex models could lead to overfitting. Thus, it is necessary to obtain more data before seeking to apply these advanced neural models as classifiers or regressors. In light of this and considering that obtaining new data is already part of our proposed future work, another line of work could be to train more powerful predictive models on an extended data set.

ENDNOTES

[1] http://www.uma.es/eees/

[2] http://www.ucn.cl/sobre-ucn/somos-ucn/proyecto-educativo-ucn/

ORCID

Víctor Flores http://orcid.org/0000-0002-6995-9434

REFERENCES

1. A. Aritzeta and S. Ayestarán, *Aplicabilidad de la teoría de los Roles de equipo de Belbin: Un estudio longitudinal comparativo con equipos de trabajo*, Revista de Psicología general y aplicada **56** (2003), 61–75.
2. P. Baepler and C. J. Murdoch, *Academic analytics and data mining in higher education*, Int. J. Scholarsh. Teach. Learn. **4** (2010) 17.
3. M. Belbin, *Management teams: Why they succeed or fail*, 3rd ed., Elsevier, London, England, 2010.
4. M. Belbin, *Team roles at work*, 2nd ed., Routledge, New York, USA, 2012.
5. R. Berenson, G. Boyles, and A. Weaver, *Emotional intelligence as a predictor of success in online learning*, Int. Rev. Res. Open Distrib. Learn. **9** (2008), 1–17.
6. L. Breiman, *Random forests*, Mach. learn. **45** (2001), 5–32.
7. A. J. Dubrin, *Human relations: Interpersonal job-oriented skills*, Pearson Higher Ed, 2014.
8. M. Fernández-Delgado et al., *Do we need hundreds of classifiers to solve real world classification problems*?, J. Mach. Learn. Res. **15** (2014) 3133–3181.
9. V. Flores and J. Gómez, *Applying active methodologies for teaching software engineering in computer engineering*, IEEE Revista Iberoamericana de Tecnologías del Aprendizaje **12** (2017), 182–190.
10. V. Flores, P. Lara, and M.C. Gaya, 2013, Project Based Engineering School: una escuela conectada con un mundo profesional sostenible. X Jornadas Internacionales de innovación universitaria, Villaviciosa de Odón, España.
11. J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning* (**Vol. 1**), Springer, Berlin, Springer series in statistics, 2001.
12. S. Geisser, *Predictive inference: An introduction*, J. R. Stat. Soc. Ser. A Stat. Soc. **158** (1995), 185.
13. K. S. Harris and G. E. Rogers, *Soft skills in the technology education classroom: What do students need*?, Technol. Teach. **68** (2008), 19–25.
14. T. K. Ho, (1995, August). Random decision forests. In Document analysis and recognition, 1995., proceedings of the third international conference on (Vol. **1**, pp. 278–282). IEEE.
15. T. K. Ho, *The random subspace method for constructing decision forests*, IEEE Trans. Pattern Anal. Mach. Intell. **20** (1998), 832–844.
16. S. T. Jishan et al., *Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique*, Decis. Analyt. **2** (2015), 1.
17. M. L. Laskey and C. J. Hetzel, Self-Regulated Learning, Metacognition, and Soft Skills: The 21st Century Learner. Online submission, 2010.
18. G. Mendez et al., *Factors associated with persistence in science and engineering majors: An exploratory study using classification trees and random forests*, J. Eng. Educ. **97** (2008), 57–70.
19. T. Mishra, D. Kumar, and S. Gupta, (2014, February). Mining students' data for prediction performance. In Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on (pp. 255–262). IEEE.
20. S. Natek and M. Zwilling, *Student data mining solution-knowledge management system related to higher education institutions*, Expert Syst. Appl. **41** (2014), 6400–6407.
21. F. Pedregosa et al., *Scikit-learn: Machine learning in python*, J. Mach. Learn. Res. **12** (2011), 2825–2830.
22. V. Ramesh, P. Parkavi, and K. Ramar, *Predicting student performance: A statistical and data mining approach*, Int. J. Comp. Appl. **63** (2013), 35–39.
23. A. Sabino et al., 2005, Guía para el trabajo en equipo. Universidad del País Vasco / Euskal Herriko Unibertsitatea. Noviembre de 2005.

24. C. Sammut and G. I. Webb, *Encyclopedia of machine learning*, Springer Science & Business Media, New York, USA, 2011.
25. A. M. Shahiri and W. Husain, *A review on predicting student's performance using data mining techniques*, Procedia Comput. Sci. **72** (2015), 414–422.
26. V. C. Smith, A. Lange, and D. R. Huston, *Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses*, J. Asynch. Learn. Netw. **16** (2012), 51–61.
27. A. Van Barneveld, K. E. Arnold, and J. P. Campbell, *Analytics in higher education: Establishing a common language*, EDUCAUSE Learn. Initiat. **1** (2012), 1–11.

L. Gutiérrez is a current student of Computer and Informatics Engineering at Universidad Católica del Norte. He has published scientific articles on the fields of machine learning, natural language processing, and statistical learning in sentiment analysis. His research interests include machine learning, text mining, information retrieval, and sentiment analysis.

V. Flores received the degree in Systems Engineering from Univ. "Santiago Mariño" (Caracas-Venezuela) in 2000. He received both MSc on Software Engineering (2003) and PhD in Computer Science (Artificial Intelligence - 2011) from "Universidad Politécnica de Madrid" (Madrid-Spain). His PhD thesis was under the advisement of professor Martín Molina. He is currently professor & researcher in the Computer & Systems Engineering from the "Universidad Católica del Norte," Antofagasta-Chile. His research interests include Intelligent Systems, Artificial Intelligence Techniques, and innovative techniques on Active Learning. His work has been documented in more than 15 papers, from 2006 to 2018, in the fields of Expert Systems, Artificial Intelligence, Active-Learning, Project Management, and Education.

B. Keith received his Bachelor's Degree in Engineering Science and a degree in Civil Engineering in Computing and Informatics in 2016 from the Catholic University of North, Chile. He also earned a Bachelor's Degree in Mathematics and a Master's Degree in Informatics Engineering from the Catholic University of North in 2017. He is currently an academic in the Department of Computing & Systems Engineering at the Catholic University of North. His areas of research include natural language processing, text mining, sentiment analysis, machine learning, and data science applications in general.

A. Quelopana received his Bachelor's Degree in Engineering Science and a degree in Civil Engineering in Computing and Informatics from the Catholic University of North, Chile. He also earned a Master's Degree in Informatics Engineering from the Catholic University of North in 2008 and he earned a Master's Degree in Engineering Project Management at The University of Melbourne, Australia. He is currently an academic in the Department of Computing & Systems Engineering at the Catholic University of North. His areas of research include project management, Information Systems, and Engineering education.

How to cite this article: Gutiérrez L, Flores V, Keith B, Quelopana A. Using the Belbin method and models for predicting the academic performance of engineering students. *Comput Appl Eng Educ*. 2019;1–10. https://doi.org/10.1002/cae.22092