A framework for predicting academic orientation using supervised machine learning

Hicham El Mrabet1,2 · Abdelaziz Ait Moussa2

Received: 13 July 2021 / Accepted: 11 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
School guidance is declared an integral part of the education and training process, as it accompanies students in their educational and professional choices. Accordingly, the current situation in light of the Covid-19 epidemic requires a reconsideration of school guidance together with the methods of accompanying the student to choose the field that suits his/her personality, knowledge qualifications, perceptual and intellectual skills in order to achieve an excellent educational level that enables the learner to work in future professions. The current study aims to predict a student's potential and provide support for academic guidance. This paper emphasizes the importance of supervised machine learning and classification algorithms to predict the personality type based on student traits. Based on the information gathered, the results of this study indicate that it contributes significantly to providing a comprehensive approach to support academic self-orientation.

Keywords Machine learning · Classification algorithms · Smart school guidance

1 Introduction
Information and communication technology (ICT) has become a strategic choice in tomorrow's school project, and one of its most essential pillars because controlling these technologies is one of the most important ways to prepare young people to face the challenges of the future. ICT in education provides more outstanding performance, visual control, better perception, and faster learning (Delić-Zimić and Gadžo 2018). In the academic journey, the student needs a particular support from specialists in media and guidance in order to be able to get a comprehensive idea of future academic prospects and job opportunities associated with future academic selection based on his/her capabilities, qualifications, desires, and personality. The main barriers faced by student resides in making the right decision about what academic path to follow. Mostly, the guidance choices fail when academic choices are based on unsupervised foundations and information or emanate from the desires of the surroundings and the parents' aspirations, which do not necessarily reflect the student's qualifications. An up-and-coming technology to achieve this objective is the use of artificial intelligence, the internet of things (El Mrabet et al. 2021a), machine learning, and data mining (Memiş et al. 2022a). The use of machine learning in an educational environment is more common in smart learning, predicting student performance, and predicting academic orientation. Machine learning applications related to predicting academic careers are increasing rapidly and can profoundly influence the field of education. It is able to track student interest, aptitudes and analyze student behavior and reactions (Asthana and Hazela 2020). There are many parameters to be taken into account to consider regarding school guidance, such as students' personalities. We can use student traits obtained from teachers in each subject to create the groups of students classified into personalities. The ability to predict a student's personality is crucial in determining their academic path and vocational career. In this paper, we will focus on supervised machine learning to classify students according to the six personalities presented by Holland's theory (Holland 1997).
2 Background

Artificial intelligence proved its efficiency in solving data science problems in all fields, and education, in turn, has had its share in this technological revolution. As a result, wide variety of machine learning algorithms from different paradigms are used successfully in education. In this section, we will review the fundamentals and algorithms of supervised machine learning and highlight specific learning and optimization approaches.

2.1 Machine learning

Machine learning (ML) is a scientific study of the algorithms that computer systems use to efficiently perform a specific task without using multiple explicit instructions (Ihya et al. 2020); it indicates how computers can learn or improve their performance data. Machine learning algorithms adopt a mathematical model that relies on data samples, called “training data,” to make predictions or decisions without being explicitly programmed to perform a specific task.

Several researchers have developed relevant classification algorithms to make the appropriate decision (Memiş et al. 2021). There are three crucial types of machine learning algorithms, as shown in Fig. 1.

2.2 Supervised machine learning

Supervised machine learning is the most popular for performing machine learning operations where algorithms rely on existing labeled data to classify new data. As the learning period progresses over time, the algorithm will identify the relationships between the predefined classes. This trained algorithm is then fed onto the unlabeled test dataset to categorize them into similar groups (Uddin et al. 2019). Among the supervised learning algorithms that we will be deal with in more detail in the following sections, we cite logistic regression, KNN, decision tree, Naive Bayes classifier, neural network supervised, SVM. Supervised learning methods are commonly leveraged in different fields to solve two main categories of problems: regression and classification (El Mrabet et al. 2021b).

2.3 Classification algorithms

In machine learning, classification is one of the essential aspects of supervised learning, in which the computer system learns from the input data and then uses this learning to classify new observations. The field of data classification is of increasing importance due to the unpredictability, large amount, and complexity of real-world data (Erkan 2020). the goal of this study is to independently determine learners’ personalities according to the traits (input data) of each student. Knowledge discovery and data mining use various classification methods and techniques. Classification algorithms generally contain two phases:

- Learning phase: In this phase, a classification model is built from the training data. The classification algorithm leverages the labeled input data to adjust the classification model by reducing errors that occurred from the learning.
- Test phase: In this phase, the built-in model is used to assign a label to an unlabeled test instance based on the model obtained in the first phase (Learning phase).

To improve the classification performance, the concept of fuzzy sets (Memiş et al. 2022b) is used, listed among the widespread mathematical tools defined to deal with uncertainty. In the following subsections, we will discuss the five major classification algorithms (Logistic regression, KNN, Decision tree, Naive Bayes, SVM), and also present some of their characteristics and how they work in order to choose
the best performing algorithm for decision making in our smart guidance system.

2.3.1 Logistic regression

It is a robust and well-established mathematical model of supervised classification, widely used in statistics to guess the posterior probability of an event based on a set of inputs (Uddin et al. 2019). Logistic regression works with binary data, which means two possible ways to find the probability of an event occurring (P = 1) or not (P = 0)—taking the example of school guidance by considering only two classes: Good orientation P = 1 and poor orientation P = 0.

2.3.2 K nearest neighbors

The k nearest neighbors (KNN) algorithm is one of the simplest, most basic, and oldest classification algorithms that store all available cases to classify new cases based on a measure of similarity (Pandey et al. 2019). It is a non-parametric classification system that completely circumvents the problem of probability densities. The 'K' represents the number of nearest neighbors considered to take the vote; this means that the decision is made by examining the labels of the K nearest samples (Jiang et al. 2012). Figure 2 illustrates how the KNN works to classify a new object. For example, for K = 3, the new object (triangle) is classified with the category "star." However, it was classified with "square" when K = 7.

2.3.3 Decision tree

The decision tree (DT) is one of the most popular classification methods, and the oldest, they appeared in the 1960s in the fields of research in psychology and sociology (Kasperczuk and Dardzinska 2018). It is represented in an intuitive tree format to classify data elements in a tree structure. The nodes of a decision tree normally have several levels where the first is called the root node (containing all the observations), the second level contains a series of branches having at least one child whose intersections produce us nodes which are called internal nodes (represent tests on input attributes). Based on the test result, the classification algorithm branches to the appropriate child nodes until it reaches the leaf node. These leaf nodes correspond to the decision results, i.e., the classes to be predicted (Uddin et al. 2019).

The general idea of a decision tree is to build a tree T from a set of observations S, as shown in Fig. 3. If all our observations S belong to a class C, then the node is considered as a leaf node and receives label C; otherwise, the algorithm selects the following most informative attribute and builds sub-trees until the last criterion is met (Khelfaoui and Sedkaoui 2020). The first step is to define the most informative attribute, generally based on the entropy method, which measures the homogeneity of a sample. The second step is to calculate the information gain that measures the relative change in entropy for the independent attribute.

2.3.4 Naive Bayes

The Naive Bayes algorithm is a classification technique based on Bayes' theorem of conditional probabilities. This theorem describes the probability of an event based on prior knowledge of the conditions associated with that event, and it states that if we have a hypothesis H and a proof E. It is a reasonably intuitive, simple, and robust algorithm for predictive modeling whose main objective is to treat each feature independently (Sedkaoui 2018).

Figure 4 illustrates a "black" triangle that represents the new sample instance which must be classified either in the "star" class or in the "square" class according to the class which achieves the most significant posterior probability. To classify the "Triangle" object, we must draw a circle around this object which includes several adjacent points.

2.3.5 Support vector machine

The support vector machine (SVM) is another simple algorithm that produces precision results with less computing...
power (Cristianini and Shawe-Taylor 2000). It is among the best-known algorithms that rely on linear or nonlinear separation data (classes) using a hyperplane (actually a line) while maximizing the marginal distance for the two classes and minimizing classification errors. As shown in Fig. 5, the marginal distance for a class is the distance between the hyperplane and the nearest instance called support vectors.

### 3 Related works

Machine learning provides a great advantage in predicting performance and organizing the field of education (Velakanti and Mathur 2020). This new paradigm in the field of education is very beneficial in representing student learning to predict the student's future behavior, such as knowledge, learning behavior, and cognitive abilities. In simple terms, machine learning technology performs analyzes based on students' data and makes the decision-making process automatic and consolidated (Asthana and Hazela 2019). The application of machine learning in education environment takes many forms to improve the teaching and learning process:

- Predicting career paths.
- Predicting student performance.
- Precise grading.
- Predicting academic guidance.
- Increasing efficiency.
- Personalization in the classroom.
- Personalized learning path.
- Personalization and choice behavior.

Several researchers focus on applying machine learning techniques and data mining to make a conscious decision in the educational field (Velakanti and Mathur 2020; Asthana and Hazela 2019; Gray and Perkins 2019; Chung and Lee 2019). Shatnawi et al. (2014) conducted a study on a smart academic advisory system called (SAAS) using data mining techniques to improve students’ performance by proposing courses that meet their current needs. The system successfully creates a list of association rules that guide specific students to choose courses that similar students enroll in. Oskouei and Askari (2014) established predictive models for high school student's performance based on several classification and prediction algorithms (Naive Bayes, C4.5, random forest, neural networks) to improve the accuracy of predicting students’ academic results. The results showed that gender, family, and parental education level influence students’ academic performance. In Zahour et al. (2020), the authors concentrate on Chatbot application as an online guidance agent to support undergraduate and graduate students to make their decisions on the choice of academic and career paths. The proposed approach focuses on students' personalities to define the predominant personality type and get a suitable career path. In El Mrabet and Moussa (2019), the writers suggest a smart school guidance system called "ETC guidance system," its role is to reduce failure academic guidance and create an environment conducive to the successful adaptation of school guidance according to students’ cognitive trends and soft skills using the internet of things. The system offers an innovative solution to guide students in smart cities to choose a conscious and promising career path. Sacin et al. (2009) propose a recommendation system based on data mining techniques to help students make a suitable decision on their academic journey. This system provides support for the student to better choose the number of courses to enroll in, based on students' experience, especially of previous students with similar academic achievements. Castellano et al. (2008) proposed collaborative recommender systems that help people in the academic orientation field in order to support counselors helping high school students to make decisions about their academic guidance. The system uses students'
grades as input data to suggest their academic possibilities by providing multiple recommendations.

4 Contribution

The current research aims to use students’ traits as a metric measure to reveal the relationship between personality behaviors and academic guidance. The main theories of personality recognition provide career orientation professionals around the world with a set of techniques and concepts they can use to guide students towards a suitable career. These theories are based on various personality models, such as big five factor personality (De Raad and Mlačić 2015), Myers Briggs type indicator (King and Mason 2020), vocational personalities in work environment (van Vianen, 2018), and DiSC personality assessment (Utami et al. 2019). In addition, the guidance system is partly designed on the theory of professional decision-making. The theory of professional decision-making is one of the most important theories that discuss the conditions for career success, methods of job selection, and the factors affecting the most appropriate choice of a professional field.

The main idea of the theory refers to the necessity of individuals using their dominant and distinct characteristics in the search for a profession that suits them. The theory also suggests that people who choose to work in an environment like their personality type are more likely to succeed and achieve a high job satisfaction rate. Holland’s theory indicates six basic types of work environments, as shown in Table 1. This contribution aims mainly to develop a robust system to predict the academic orientation path of the student from the analysis of personality traits. Therefore, this work comes to tackle the limitations of previous research works regarding academic orientation, where most of the papers cited in the related works section followed a simple method based on student grades, gender, personality as a metric of orientation; furthermore, they did not go through a qualitative approach based on an ontological paradigm of applying student traits studies. The main contribution of this work is the ability to predict the student’s academic path by collecting a learner’s personality traits from several teachers and continuously processing those trait evaluations using classification algorithms. We used for that purpose an online learner paradigm that improves the classification models with continuous-increasing of student trait evaluation data.

5 Research methodology

The main objective of this study is to advance knowledge in the field of vocational guidance to discover patterns to suggest the right academic guidance. More precisely, this research has two specific objectives. The first is to identify the most associated traits with a student’s personality. The second objective is to determine the traits of individuals who have high interests in a specific professional field.

5.1 Personality classification

In this research, the learner personality classification focuses on supervised machine learning. Figure 6 illustrates the methodology used in this research, which is based on four algorithms: decision tree, logistic regression, KNN, and SVM. We used the following ML tools to test these classification algorithms, namely Python, Anaconda, and Jupyter Notebook. The easiest way to program with Python is to install the Anaconda distribution, which bundles several preinstalled tools and libraries for scientific computing and data science. Among the tools offered by Anaconda, an interactive environment called Jupyter. It is a web application that allows you to share code and run it in the same user interface.

5.2 Data collection

The system uses actual data from the registration pool of qualifying high school learners. The data collected in this test represents a database of 350 samples. Before giving a reading on the result of the performance, which was deduced based on the different graphs of ROC sensitivity/specificity curves “receiver operating characteristics,” it is important to describe the framework adopted in this performance evaluation:

- Split training data and test data have been adopted in this performance test by removing 30% of the complete data set for test data, and 70% of the data is dedicated to training data. It is very important that the training data is more important than the test data to be sure that the models are trained with sufficient data variability. In contrast, test data is used to assess effectiveness (actual performance).
- The data-set size went up to 50,000 observations, with only 350 real observations taken from the field; the rest is generated from an extrapolation technique with an injection of a few noise observations.
5.3 The predictive model using the decision tree

The performance evaluation of the decision tree (DT) classification model for the six classes (six personalities for the classification) was also using the ROC curve. Figure 7 illustrates the result of the graph obtained.

The adjustment and regularization parameters of the DT model are chosen after establishing the "cross-validation” technique. The following parameters represent the configuration adopted to evaluate our DT classification model:

- Maximum depth: represents the maximum depth to be considered for creating the decision tree. This parameter significantly impacts the algorithm’s convergence because if the depth is very small, the algorithm may
A framework for predicting academic orientation using supervised machine learning

stop before it converges. In our case, this parameter is configured to a maximum of 100.

– Minimum samples for node division: This parameter is configured with the default value of 2. It represents the number of observations required in a node to divide it.

As shown in Fig. 7, the graph represents the six ROC curves of each personality class produced during the evaluation of the decision tree model. The six ROC curves are pretty close, where the AUC is between 0.77 and 0.87 with an average of 0.80, which gives a fairly good classification performance reading. Yet the difference obtained between the six personality classes is judged by having an unbalanced data distribution or perhaps by a probable introduction of outliers of the traits during the evaluation by some teachers. For example, in the ROC graph of DT, it turns out that when the TPR is between 0 and 0.6, the FPR is almost zero, so if TPR goes up from 0.6, the FPR is linearly increasing with the TPR. This means that the more the model gives correct predictions for a personality class, the more it even predicts the personality class in the classification of other personality classes that are judged to be incorrect.

5.4 The predictive model using the logistic regression

The performance evaluation of the "logistic regression" (LR) model for the six personality classes was using the ROC curve. Figure 8 illustrates the result of the graph obtained. The model's fit and regularization parameters are chosen after several tests via training and test data variation using techniques such as "cross-validation." The following adjustment and regularization parameters represent the configuration adopted to evaluate our RL classification model:

– Maximum iterations: represents the maximum number of iterations-epochs for the differentiable optimization algorithm (gradient descent) to end. In our case, this parameter is configured to a maximum of 5000 iterations.

– Learning rate: represents the amount with which the cost function (error function) decreases. Technically speaking, this is the pace at which the optimization algorithm moves. In our case, the learning rate is set at 0.2.

As shown in Fig. 8, the graph of the ROC curve reflects six ROC curves, each corresponding to a personality class. The six ROC curves are quite close, where the AUC “Area Under Curve” is between 0.66 and 0.78, with an average of 0.73. Yet personality class 1 had the lowest AUC value, which means that this personality is ranked lower compared to other personalities. This can be judged by having:

– An unbalanced distribution of data between personality 1 and the other personalities: in the data-set used, the density of observations for personality 1 is very small compared to the other personalities.

Fig. 6 Systematic approach to predicting personality

Fig. 7 ROC curve of the decision tree model

Fig. 8 ROC curve of the logistic regression model
– Possible, there were bad evaluations obtained by some teachers (outliers of traits) for personality 1.

Seeing that the problem we are trying to solve in our context is obviously a multiple classification (six personality classes) and not a binary classification, so this implies finding a decision threshold that leads to maximizing the TPR "true positive rate" and at the same time to minimize the FPR "false positive rate" of the average ROC curve for the six personality classes, in our case we see that with TPR of 0.7~ and FPR of 0.34~ we could choose a suitable decision threshold for the RL algorithm.

5.5 The predictive model using the KNN

The performance evaluation of the KNN classification model for the six personality classes was also using the ROC curve. Figure 9 illustrates the result of the graph obtained. The “K” regularization parameter of the KNN model is chosen after performing the “cross-validation” technique. In our case, the parameter K is configured with the value 17; this means that the KNN algorithm will be based on the 17 observations closest to the new value to be predicted to perform the classification. Just a reminder, the KNN does not have a training step, but on the contrary, it goes directly to the classification step for each new entry. This means that the KNN algorithm is always expensive in terms of time-consuming when predicting a new value. However, as illustrated in Fig. 9, the graph shows the six produced ROC curves corresponding to each personality class, its curves are quite very close, or the AUC is between 0.61 and 0.75 with an average of 0.65, which gives a classification performance reading pretty average. Furthermore, as the ROC graph illustrates, it turns out that when the TPR is between 0 and 0.3, the FPR is almost zero, so if TPR goes up by 0.3, the FPR is linearly increasing with the TPR.

5.6 The predictive model using the SVM

Figure 10 shows the performance evaluation of the "SVM" classification model for the six personality classes. After establishing the "cross-validation" technique, the model’s adjustment and regularization parameters are chosen.

The following parameters represent the final configuration adopted to evaluate our SVM classification model:

– Maximum iteration: the maximum number of iteration-epochs for the SVM optimization algorithm to complete is configured to 10,000 iterations.
– Learning rate: In our assessment test, the learning rate is set at 0.08.

As the SVM ROC graph illustrates, the six ROC curves for each personality class are a bit scattered, where the AUC is between 0.68 and 0.82 with an average of 0.72, which gives a reading of fairly average classification performance. In our case, the SVM model took enough time to converge due to the learning rate being too small (set to 0.08); this implies more iterations added by the optimization algorithm. As the ROC graph illustrates, it turns out that when the TPR...
is between 0 and 0.5, the FPR is almost constant in the value 0, so if TPR goes up by 0.5, the FPR is linearly increasing with the TPR.

### 6 Interpretation summary

The result obtained from the above ROC curves and also time-consuming captured during our benchmark test for the different algorithms: KNN, LR, DT, and SVM are summarized in Tables 2 and 3. The result shows that the average performance of all classes for different models are not above 80%, where the DT model comes in the first place with an AUC average of 0.8, it follows the LR model with an average AUC of 0.73, after that, it comes the SVM model with an average AUC of 0.72 and finally the KNN with only an average AUC of 0.65. In terms of time-consuming, Table 2 shows that the KNN and SVM are very expensive compared to LR and DT. For KNN, the K has chosen, which is K = 17, is a reason for the huge time-consuming that could be reduced during the ever-increasing of new data (new teachers’ evaluations).

The result cannot be considered sufficient for now to confirm the model maturity, and we can explain that with the following reasons:

- The distribution of the dataset used is imbalanced, from the 350 student trait evaluations, some traits may not be significantly present, and that could then the others may dominate the fitting model process, thus, may cause the model to be unable to learn from the other traits with lower variance.

In our study for the smart school guidance, two points are essential to judge the maturity of our approach:

- Our approach requires following an online-learning strategy that permits the model chosen to auto-improve the performance in real-time while introducing new teachers’ evaluations. Hence, selecting the suitable model will require taking into account the time consumption of the model during the prediction stage; for example, the KNN may not be very suitable in this context.

- We initiated our approach by predicting student personalities based on the trait evaluation established by the teachers; then, it comes into play the second part of this approach by predicting the professional careers of each student based on its predicted personality.

### 7 Conclusion

This research offers a promising perspective for new learning environments for ensuring quality education. On the other hand, this research introduces the concept of smart school based on integrating a career guidance system so that the awareness of guidance of students in innovative institutions is beneficial in providing informed career choices. The reflections presented in this research fall within the framework of designing an intelligent orientation environment that depends on the personality type. Indeed, to achieve this goal, we were interested in machine learning techniques and, in particular, in classification algorithms. The approach adopted consists of measuring the performance of classification algorithms to predict the personality of each learner. We carried out a method of predicting the school path based on the learner’s traits. Several supervised machine learning techniques are evaluated based on their predictive ability, including KNN, SVM, decision tree, and logistic regression, and decision tree and logistic regression are shown to offer good prediction accuracy. Specifically, the decision tree has demonstrated up to 0.80 successes in terms of performance. The present findings will be an essential step for future research on school guidance. From the preliminary results obtained, we noted that it is very difficult to make
a relevant choice among the classification algorithms since the actual number of samples is very small, so future studies should include more data on educational guidance. It is also preferable to build a database of educational guidance over several years. This requires the development of central data acquisition, monitoring, and storage system. Using these databases to train classification algorithms can improve the prediction of a learner’s academic path. Our approach potentially impacts the guidance system, and combinations of these methods can better design an automatic guidance system. Future works aim to predict the learner’s suitable academic and professional sector according to various criteria, such as personality traits, cognitive skills, cognitive abilities, and tendencies. We rely on more widely used machine learning algorithms for classification problems that we did not mention in the manuscript, especially the Neural Network algorithm.

**Funding** Not applicable.

**Declarations**

**Conflict of interest** Not applicable.

**Availability of data and material** Not applicable.

**References**

Asthana P, Hazela B (2019) Applications of machine learning in improving learning environment. Multimed Big Data Comput IoT Appl. https://doi.org/10.1007/978-981-13-8759-3_16

Asthana P, Hazela B (2020) Applications of machine learning in improving learning environment. In: Tanwar S, Tyagi S, Kumar N (eds) Multimedia big data computing for IoT applications. Intelligent systems reference library, vol 163. Springer, Singapore. https://doi.org/10.1007/978-981-13-8759-3_16

Castellano EJ, Martínez L, Sánchez PJ (2008) OrieB, a linguistic CRS for supporting decision making in academic orientation. In: World scientific proceedings series on computer engineering and information science computational intelligence in decision and control, Chapter 2, pp 841–846. https://doi.org/10.1142/9789812799470_0138

Chung JY, Lee S (2019) Dropout early warning systems for high school students using machine learning. Child Youth Serv Rev 96:346–353. https://doi.org/10.1016/j.childyouth.2018.11.030

Cristiani N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge

De Raad B, Mlačić B (2015) Big five factor model, theory and structure. In: International encyclopedia of the social & behavioral sciences, pp 559–566. https://doi.org/10.1016/b978-0-08-097086-8.25066-6

Delić-Zimić A, Gadžo N (2018) Implementation of ICT in education. In: Hadžikadić M, Avidaković S (eds) Advanced technologies, systems, and applications II. IAT 2017. Lecture notes in networks and systems, vol 28. Springer, Cham. https://doi.org/10.1007/978-3-319-71321-2_18

El Mrabet H, Moussa AA (2019) Smart school guidance and vocational guidance system through the internet of things. In: Proceedings of the 2nd international conference on networking, information systems & security—NISS19. https://doi.org/10.1145/3320326.3320404

El Mrabet H, Ait Moussa A (2021a) IoT-school guidance: A holistic approach to vocational self-awareness & career path. Educ Inf Technol 26:5439–5456. https://doi.org/10.1007/s10639-021-10548-6

El Mrabet MA, Elkakkaoû K, Faize A (2021b) Supervised machine learning: a survey. In: 4th international conference on advanced communication technologies and networking (CommNet). https://doi.org/10.1109/commnet52204.202.9641998

Erkan U (2020) A precise and stable machine learning algorithm: eigenvalue classification (EigenClass). Neural Comput Appl 33(10):5381–5392. https://doi.org/10.1007/s00521-020-03543-2

Gray CC, Perkins D (2019) Utilizing early engagement and machine learning to predict student outcomes. Comput Educ 131:22–32. https://doi.org/10.1016/j.compedu.2018.12.006

Holland JL (1997). Making vocational choices: a theory of vocational personalities and work environments, 3rd edn. Psychological Assessment Resources

Ihya R, Aitdaoud M, Namir A, Guerres FZ, Haddani H (2020) Using machine learning algorithms to predict the e-orientation systems acceptancy. In: Ben Ahmed M, Boudhir A, Santos D, El Aroussi M, Karas I (eds) Innovations in smart cities applications edition 3, Lecture notes in intelligent transportation and infrastructure. Springer, Cham, pp 117–130. https://doi.org/10.1007/978-3-030-37629-1_10

Jiang S, Pang G, Wu M, Kuang L (2012) An improved K-nearest-neighbor algorithm for text categorization. Expert Syst Appl 39(1):1503–1509. https://doi.org/10.1016/j.eswa.2011.08.040

Kaspersczuk A, Dardzinska A (2018) Comprehensive review of classification algorithms for medical information system. In: Dang T, King J, Wagner R, Thoi N, Takizawa M (eds) Future data and security engineering. FDSE 2018. Lecture notes in computer science, vol 11251. Springer, Cham, pp. 299–309. https://doi.org/10.1007/978-3-030-31923-3_23

Khelfaoui M, Sedkaoui S (2020) Sharing economy and big data analytics, 1st edn. Wiley, Hoboken, pp 171–194

King SP, Mason BA (2020) Myers–Briggs type indicator. In: The Wiley encyclopedia of personality and individual differences, pp 315–319. https://doi.org/10.1002/9781119547167.ch123

Memiş S, Enginoğlu S, Erkan U (2021) Numerical data classification via distance-based similarity measures of fuzzy parameterized fuzzy soft matrices. IEEE Access 9:88583–88601. https://doi.org/10.1109/ACCESS.2021.3089849

Memiş S, Enginoğlu S, Erkan U (2022a) A classification method in machine learning based on soft decision-making via fuzzy parameterized fuzzy soft matrices. Soft Comput 26(3):1165–1180. https://doi.org/10.1007/s00500-021-06553-z

Memiş S, Enginoğlu S, Erkan U (2022b) A new classification method using soft decision-making based on an aggregation operator of fuzzy parameterized fuzzy soft matrices. Turk J Electr Eng Comput Sci. https://doi.org/10.3906/elk-2106-28

Oskouei RJ, Askari M (2014) Predicting academic performance with applying data mining techniques (generalizing the results of two
different case studies). Comput Eng Appl J 3(2):79–88. https://doi.org/10.18495/comengapp.v3i2.81
Pandey S, Sharma V, Agrawal G (2019) Modification of KNN Algorithm. Int J Eng Comput Sci 8(11):24869–24877. https://doi.org/10.18535/ijecs/v8i11.4383
Sacin CV, Agapito JB, Shafti L, Ortizosa A (2009) Recommendation in higher education using data mining techniques. In: Paper presented at the educational data mining 2009
Sedkaoui S (2018) Big data analytics for entrepreneurial success: emerging research and opportunities. IGI Global, New York
Shatnawi R, Althebyan Q, Ghalib B, AlMaolegi M (2014) Building a smart academic advising system using association rule mining. http://arxiv.org/abs/1407.1807
Uddin S, Khan A, Hossain M et al (2019) Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 19:281. https://doi.org/10.1186/s12911-019-1004-8
Utami E, Hartanto AD, Adi S, Oyong I, Raharjo S (2019) Profiling analysis of DISC personality traits based on Twitter posts in Bahasa Indonesia. J King Saud Univ Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2019.10.008
Van Vianen AEM (2018) Person-environment fit: a review of its basic tenets. Annu Rev Organ Psych Organ Behav 5(1):75–101. https://doi.org/10.1146/annurev-orgpsych-032117-104702
Velakanti G, Mathur A (2020) Machine learning approach to find the abilities in a candidate for steady employment in engineering field: a literature survey. Mater Today Proc. https://doi.org/10.1016/j.matpr.2020.10.585
Zahour O, Benlahmar EH, Eddaoui A, Ouchra H, Hourrane O (2020) A system for educational and vocational guidance in Morocco: Chatbot e-orientation. Proc Comput Sci 175:554–559. https://doi.org/10.1016/j.procs.2020.07.079

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.