Towards a Smart Guidance System in CAMPUSFASO : Simulation Results

Ozias Bombiri[1], Tounwendyam F. Ouédraogo[2], Paonouor Some[3], Pasteur Poda
{ozibombe@hotmail.com[1], frederic.ouedraogo@unz.bf[2], psome@unz.bf}

Université Nazi BONI[1], Université Norbert ZONGO[2]

Abstract. Guidance is a complex and multidisciplinary field where the main goal is to help students find their suitable training pathways. The emergence of artificial intelligence has boosted many area of research. Machine learning tools have been used to improve both educational and vocational guidance system. Since 2018, the university guidance system has evolved with the establishment of an online platform named CAMPUSFASO. This platform, presented as an innovation for the guidance, has been strongly criticized. Firstly, we present the academic achievements of the first year students of Université Norbert Zongo after they are guided by CAMPUSFASO. These academic achievements show that more the student is guided at his preferential training path more he succeed. Secondly, we present a machine learning model for the guidance. Unlike CAMPUSFASO, our model uses of high school grades of the student to find the suitable training path. The model reaches auspicious results with simulated data.

Keywords: Guidance system, Machine learning, Training pathway, CAMPUSFASO, Université Norbert Zongo

1 Introduction

The guidance is an important factor in school and academic education and always necessary in the vocational path. Performing a suitable guidance has always been a concern of the main actors of the education. Many contributions have been done by researchers but it still stays more to understand in this complex topic [1, 2, 3, 4, 5]. The machine learning boom has emerged many algorithms an tools used in various areas of research [6, 7, 8, 9]. The educational guidance is an example where machine learning is increasingly used as recent works show [5, 10, 4, 11, 12].

In [11], the authors proposed a school guidance system named "IoT-School Guidance". They idea is a new approach that includes implementing an adaptive orientation process using smart technologies to guide pre-tertiary students choose and follow the best professional career generated automatically by the IoT-school guidance system.

In this work, we deal with the guidance of new baccalaureate graduates to the training pathways of the universities of Burkina Faso. The guidance process made at the high school end to

JRI 2021, November 11-13, Bobo-Dioulasso, Burkina Faso

DOI 10.4108/eai.11-11-2021.2317973

allow new baccalaureates to find their suitable pathways training has evolved with the opening of new universities. Some universities like Université Norbert ZONGO have implemented their own guidance process, different from that of Université Joseph KI-ZERBO which is the first university of Burkina Faso.

With the aim of computerizing the guidance system, the minister of higher education of Burkina Faso has launched in 2018 the online platform called CAMPUSFASO to perform the guidance of baccalaureates in all universities. The business core of CAMPUSFASO is based mainly on the baccalaureate marks and the available training pathways. However only the baccalaureate mark in a subject cannot judge the student's level in that subject. Therefore the Guidance process has some shortcomings, such as assignments in pathways that are inconsistent with the student's ambitions. Our contribution in this paper is to show how machine learning methods may improve significantly the guidance process of CAMPUSFASO. We propose a approch of guidance to taking into account the marks obtained by the student during his high school. We designed models lies on machine learning methods and the one of neural networks gives the best results. We make use of artificial data, obtained by simulation.

The reminder of this paper is organized as follow. The section 2 deals with related works. The section 3 is the background of the guidance process of public universities in Burkina Faso. The section 4 presents the academic achievement of the first students guided by CAMPUSFASO at Université Norbert ZONGO. The section 5 presents the simulated data and the design of the guidance model and the section 6 presents main results and discussion.

2 Related works

Many works have been done to improve the guidance through machine learning tools [13, 10, 14, 12, 15, 16]. In [13], the guidance system designed is to help students who have completed their high school education and must choose an appropriate career path. The system performs an online assessment of the students' skills and then an appropriate training path is selected. This contribution is similar to ours with the exception of the assessment stage where we use the student academic achievements to get the appropriate career path. This recent contribution [4, 10] propose to use the chatbots to get profile or personality through a questionnaire. Their system aims to provide guidance for undergraduate and graduate students looking for job. All these contributions show that the use of machine learning in a guidance system is an interesting area of research worth pursuing.

3 Background on guidance

After completing their high school program to obtain the baccalaureate degree, students have to pass through a guidance system to get in the public and some private universities. The guidance is done online by CAMPUSFASO platform, where future students have to apply by filling a form. The student have to provide his baccalaureate identifying number and eighteen (18) choices of pathways within all the pathways available in all the universities. This information of the applicants is send to the system that perform the guidance basis on their choices and baccalaureate scores.

The universities of Burkina Faso had their own guidance processes before the CAMPUSFASO platform that has unified to one guidance. After the baccalaureate exam, each university launches the recruitment of new students. At that moment, there were three universities including Université Norbert ZONGO which had its own guidance different from the other two. Each applicant had to pass through a team of guidance counselors. By question-answers, the counselors try to identify the applicant profile and the motivations behind his choices of the training proposals. This information allows the counselors to carry out the best guidance for the future student. In contrast, The universities of Ouagadougou (currently Université Joseph ky ZERBO) and Bobo-Dioulasso (currently Université Nazi BONI) had their own guidance processes. The applicant filled out a form where he made three choices among the proposal training pathways of the university. Future students had no interview with guidance counselors. The guidance relies only on the baccalaureate exam marks. Therefore we can say that Université Norbert ZONGO had the best guidance process because it takes into account the profile or abilities and will of applicants. CAMPUSFASO is a digital implementation of the two universities guidance system with more number of choices. It aims to bring together all guidances made by each university into a unique platform. Henceforth, every applicant has to pass by the platform to get access to the universities. He fill out a form were he has to made 18 choices among the training pathways of all universities. And it is at this level that the complaints against the platform begin. On the one hand, making 18 preferential choices seems to be too many to make a suitable guidance of the applicant. The idea behind this, is to feed the new universities with students. These new universities are open in small town far from the capital Ouagadougou where there are more services and amenities. But a student strained to make 18 choices will certainly make choices that do not match his ambitions. On the other hand, the witnesses bring out cases of studies stop because of assignments that did not go well, either because of the pathway, or because of the city of the university. In the student's preferences a choice may have a low priority but it will be retained because the process is based on the grades of the baccalaureate. Guidance at a university in a city where the student does not have a family or a guardian can also lead to cases of study stop, because university housing estates when they exist are not able to accommodate all students and many of them cannot rent a private accommodation. In the following section, we will present the results of the students assigned to the university by the CAMPUSFASO platform.

4 First year results of CAMPUSFASO at Univeristé Norbert ZONGO

CAMPUSFASO has been launched in 2018 and the first class students coming from its platform should be at their last year of the *licence*[1]. Due to the accumulated delays in performed the courses, only the results of the first two semesters are available. Table 1 shows the academic achievements by training stream of two first semesters of the first class students of CAMPUSFASO at Université Norbert ZONGO. Most of the training pathways has more than 60% of completion rate at the first and second semesters. Faced with these achievements, we are attempted to know how was the success rates in the same training pathways before CAMPUSFASO. Unfortunately we have got partial data which cannot be used in this paper.

[1] the first degree of the LMD (Licence Master Doctorat) model in effect in the universities of Burkina Faso.

Table 1: Academic achievements of the CAMPUSFASO first class students of at the Université Norbert ZONGO.

Training pathway	**# of students**	**Completion rate**	
		Semester 1	Semester 2
Geography	858	63.28%	78.88%
History and archeology	1928	70.23%	59.06%
Modern literature(French)	1719	68.35%	44.58%
Math., physics, and computer sciences	469	13.43%	55.84%
Philosophy	252	37.70%	12.69%
Psychology	111	62.16%	38.739%
Life and earth sciences	960	71.04%	47.70%
science of documentary information	148	69.59%	77.70%
Economy	1050	55.52%	68.57%
Total	7495	62.27%	56.22%

We present in the Figure 1 the success rates of students according to rank of the choice which was accepted by the guidance system of CAMPUSFASO. The number 1 are those who have been assigned to their first choice of training pathway. The number 2 are those who missed their first chose but have been accepted at their second choice, and so on up to 18. We present the first up to the seventh preferential choice because there little students assigned to their choice exceeding the 7th. The Figure 1 shows the 2 curves decreasing with the rank of the preferential choice. This means that the success is related to the rank of accepted choice. The students are more likely to fail when they miss their first choice of training pathways, and this get worse when the rank of the accepted choice get up. The conclusions made at the Université Norbert ZONGO could be observed at the others universities. Our results show that the guidance system of the CAMPUSFASO need to be improved in order to increase the completion rate. That is why we propose to make use of machine learning in the guidance system in order to make the platform smarter.

5 Design of the guidance model

In this section, we present a machine learning based approach for smart guidance of new graduates of baccalaureate diploma. The modeling is done with generated data. We generate the data with different strategies which give our model interesting preliminary results. The aim of this model study was to show that in the process of university guidance we can take into account some informations which will probably improve the results. We want the model that can predict the pathway in which the student is more likely to succeed from his grades. To achieve this, the model must be based on historical data, that is to say on the cases of guidance where students have successfully completed their university studies. It is therefore the assumption used in the generation of simulation data.

After the BEPC diploma, pupils must be guided to continue their studies in high school. The high school lasts 3 years, one year per class (*Seconde*, *Première* and *Terminale*) and the last one leads to the baccalaureate diploma. There are scientific, literary, technical and vocational streams

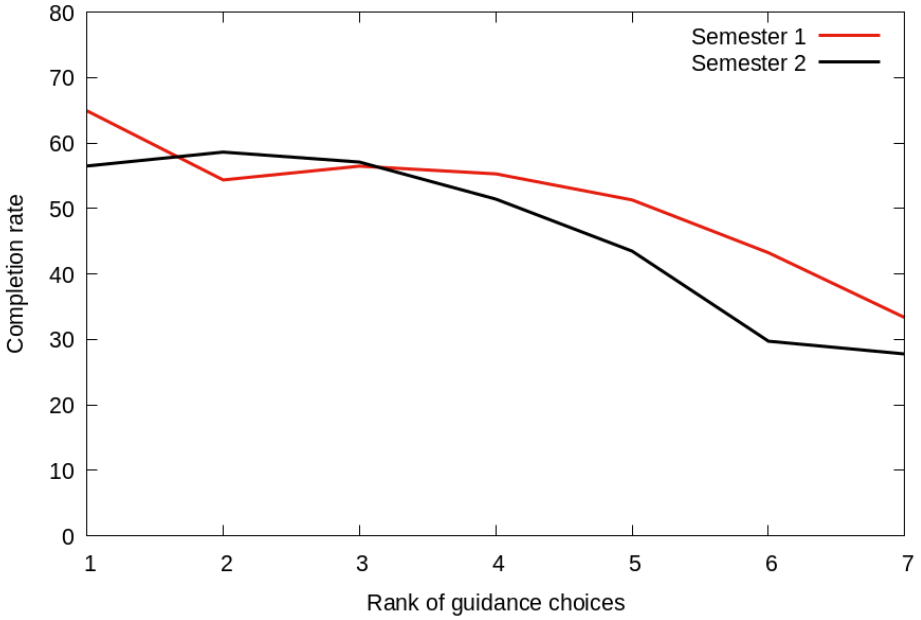

Fig. 1. Distribution of completion rate according to the rank of the choice accepted by CAMPUSFASO

where pupils are guided.

There are two fields in the science stream which lead to two types of baccalaureate, C and D. We have restricted this study to the science stream leading to the baccalaureate D. Therefore, we only consider the training pathways at the university in which the holders of the baccalaureate D of science stream can pursue their studies. Figure 2 shows these training pathways which are mathematics, physics, computer science, economy, geography and Life and earth sciences also called natural sciences.

5.1 Data generation and model fitting

We present in this part the model we devised. We show how the data is generated to fit the model. Considering our assumptions made in the section, where each new student is assigned to one and only one training pathway among the four, the guidance is a multi-class classification. Multi-class classification deals with more than 2 classes. In our case we have four classes corresponding to the four training streams: Mathematics, Physics, Computer Science, abbreviated MPCS, Life and Earth Sciences, abbreviated LES, Economy and Geography. We take into account the grades of the three years of the high school and those of the baccalaureate in the guidance model. There are eight subjects in the high school program which are mathematics, physical sciences, life and earth sciences, English, French, physical and sports education, history-geography, and philosophy. These

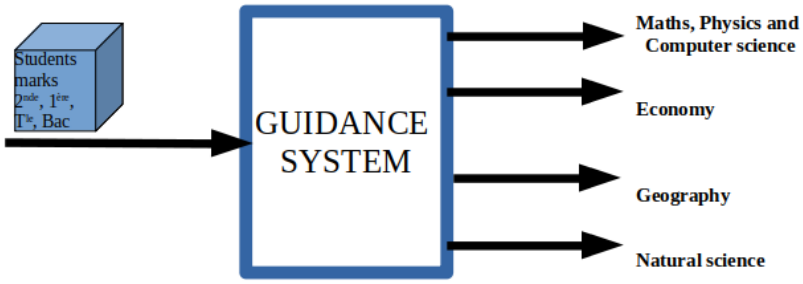

Fig. 2. Guidance system of a new graduate of baccalaureate D (science stream)

subjects are those of the baccalaureate exam. Therefore, each entry of machine learning model is 8x4 numbers(grades) that are the features and the predicted output is one training pathway, that is the label.

5.2 Dataset

We generated the data to fit the model into 2 groups, based on how the sample is labeled. The unlabeled sample, composed of 8x4 grades, is generated according to the following rules. Each grade is a decimal number, randomly chosen, between 5 to 20. Considering the coefficient of each subject, the average grades per class and the one of baccalaureate in the unlabeled sample must be equal or greater than 10. The set of grades as a matrix G is defined by:

$$\begin{cases} G = \left(g_{ij}\right)_{\substack{1\leqslant i\leqslant 8 \\ 1\leqslant j\leqslant 4}} \\ g_{ij} \in [5;20] \text{ with } g_{ij} \in \mathbb{D} \end{cases} \tag{1}$$

$$\forall j \, , \, \frac{\sum_{i=1}^{8} g_{ij}.ci}{\sum_{i=1}^{8} ci} \geqslant 10 \tag{2}$$

Obviously, the latter rule gives consistent to the generated data. It means that the baccalaureate graduate has completed the high school without repeating one class.

We label the samples in two manners. The first way is random. The second way to label is more realistic, as shown by the algorithm 1. Each sample is label under certain conditions on the grades. For instances the samples having high grades in mathematics will be label MPCS, those with high grades in life and earth sciences will be label LES, and so forth. All labels have roughly the same size, see table2.

Result: One label
Input: A sample of 32 grades ;
Initialization : ;
m_lit : average_of_literary_subjects;
m_sci : average_of_science_subjects;
if *m_sci* $\geq$ *14* **then**
  if *maths_grade* $\geq$ *life_earth_sciences_grade* **then**
    return Math. Physcis and Computer Sc. label (MPCS)
  else
    return Life and earth sciences label (LES)
  end if
else
  if *m_lit* $\geq$ *13 and m_sci* $\leq$ *m_lit* **then**
    return Geography
  else
  end if
  if *maths_grade* $\geq$ *PC_grade and PC_grade* $\leq$ *12* **then**
    return Economy
  else
  end if
  return chosen randomly label
end if

Algorithm 1: Algo : Samples labeling

Table 2: Distribution of the labeled samples. The first dataset of 5000 samples is randomly labeled and the second dataset of 5000 samples is realistically labeled.

Dataset	MPCS	LES	Economy	Geography	Total
Random labeling	1255	1219	1223	1303	5000
Realistic labeling	1434	1322	1142	1102	5000

Many machine learning methods such as support vector Machine, logistic regression, naive Bayes classifier, decision tree, random forest and neural networks have been developed and they are used in various domains. The efficiency of each method may differ from each other according to the problem being addressed. But random forest and neural networks are among the best machine learning methods. The Random Forest algorithm is one of the ensemble algorithms which combine independent weak models to have a stronger global model. It is based on the decision tree algorithm. In the Random Forest for classification, prediction is made by majority vote among the trees that make up the forest. Its advantage is that it is a simple and easy to train algorithm. The Convolutional Neural Network is a type of neural networks that suitable in processing data that has a grid form. Neural networks allow modeling with multi-dimensional data. For a student, there is a certain dependence between the grades of subjects at different levels of study(*Seconde, première, Terminale*). These grades can be viewed as serial data. Then, we can represent the grades by a table where the levels are the rows and the subjects are the columns. With this grid form representation, neural networks are suitable to process this data. All these arguments show our choice of using these two machine learning methods for the modeling of the guidance system.

We built two models of guidance with the Random Forest(RF) algorithm and the Convolutional Neural Networks (CNN). For the CNN model, we used the architecture of the AlexNet network and adapted it to our problem. AlexNet is a convolutional neural network architecture, designed by Alex Krizhevsky in collaboration with Ilya Sutskever and Geoffrey Hinton [17]. AlexNet participated in the ImageNet Large Scale Visual Recognition Challenge and achieved a top-5 error of 15.3 %, more than 10.8 percentage points lower than that of the finalist. The adaptation made in this network consists of modifying the dimensions of the input shape and the kernel size of the convolution layers. The architecture of our network is given by Figure 3.

Training and testing were performed using a cross-validation strategy with five folds (5 cross-validation). For each dataset one Random Forest model and one CNN model are trained and evaluated

6 Results and discussion

We present in this section the results of the models fitting with the two data sets, randomly and realistically labeled. The performance measurement metrics are accuracy, recall, and precision. Accuracy is the rate of good predictions on all data. The recall represents the average of the recalls for all classes. For a class, the recall corresponds to the rate of samples which are from this class and which were predicted as such. The precision is the average of the precision for all classes. For a class, the precision is the rate of correct predictions among all the samples which were predicted as belonging to this class. The Table 3 give the results obtained after fitting and testing models among these metrics.

The results in Table 3 show that the model fit with realistic dataset is significantly better than the one fit with the random dataset. In indeed, with random data, the model fails to capture information in the data structure because there is no relationship between grades and subjects. On the other hand, with the realistic data the model learns the mapping between grades and subjects. This mapping corresponds to the algorithm 1 that we used for the realistic labeling. If all the samples were

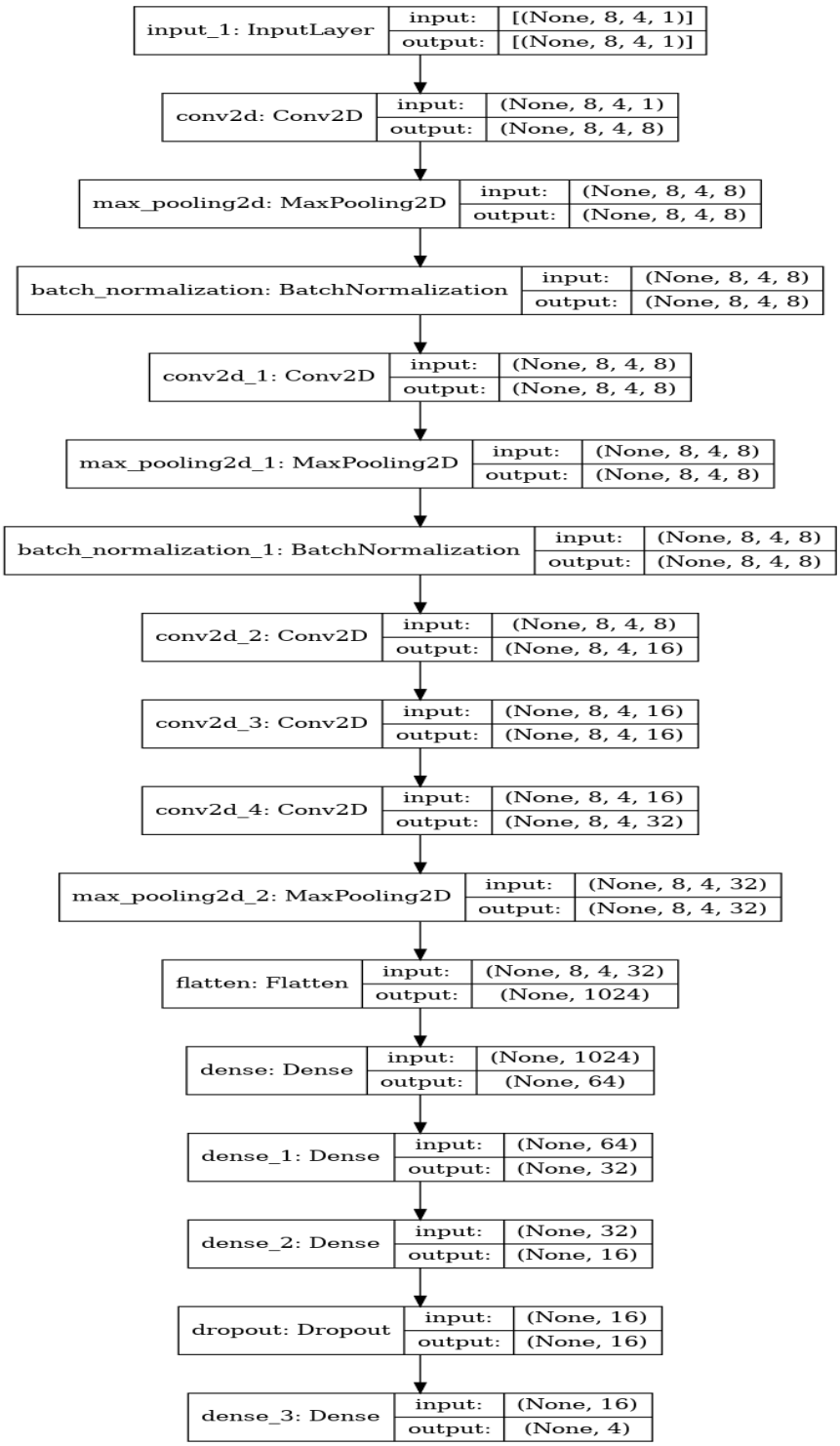

Fig. 3. Architecture of convolutional neural network used one this paper

Table 3: Results of models evaluation

Dataset	Algorithms	Accuracy (%)	Recall (%)	Precision (%)
Random labeling	RF	25.04	24.85	24.84
Random labeling	CNN	26.18	25.70	25.55
Realistic labeling	RF	61.76	60.50	61.76
Realistic labeling	CNN	79.40	78.83	79.96

deterministic then with a good model could theoretically have 100% accuracy. But we introduced a noise in the algorithm which generate the samples and some of them are still labeled randomly even in the realistic dataset. In summarize, the complexity of the labeling induces a complexity of the data structure, and therefore a variance of the model. The noise in the data used to fit the model explains the rate of efficiency achieved.

For this model simulation the objective was to show that a proof that a machine learning model could be used to capture information hidden in a person's notes. However, we do not claim that it is ready to use because it was not designed with real data. For a predictive orientation model, we want to assess each student's chances of success if he or she was oriented in a particular field. The strategy used by CAMPUSFASO also consists of calculating a mark for each student's choice. On the one hand, the advantage of our proposal is that the calculation of this mark take into account more other marks of the pupil and not only those obtained in the baccalaureate exam because the longer is the period taken, the more a pattern emerges in grades. On the other hand, the algorithm is no longer deterministic but is based on artificial intelligence to calculate the scores. By extension, other criteria may be taken into account in addition to the scores.

7 Conclusion

We presented the statistical results of the first class students of CAMPUSFASO at the Université Norbert ZONGO. The first two semesters results have shown interesting facts about the guidance system of CAMPUSFASO. Therefore, we have concluded that there is a need to improve the guidance system used. We have designed two models of machine learning, the Random Forest and Convolutional Neural Networks algorithms. The models have been fitted with simulated data as mentioned above. We obtained good results with both models. We presented the results based on artificial data because getting real data is a challenge at this step. Thus the main perspectives of this work are : obtain real data of high school grades and others relevant features to fit the models; involve other experts of guidance like psychologists in the project. The work done in this paper is the first step of a important project that the Minister of Higher Education should be involved in the next steps.

References

[1] Ryan CW, Drummond RJ, Shannon MD. Guidance Information System: An Analysis of Impact on School Counseling. The School Counselor. 1980;28(2):93-7.

[2] Shovon MHI, Haque M. Prediction of student academic performance by an application of k-means clustering algorithm. International Journal of Advanced Research in Computer Science and Software Engineering. 2012;2(7):353-5.

[3] El Haji E, Azmani A, El Harzli M. Expert system design for educational and vocational guidance, using a multi-agent system. In: 2014 International Conference on Multimedia Computing and Systems (ICMCS); 2014. p. 1018-24.

[4] Zahour O, Benlahmar E, Eddaoui A, Ouchra H, Hourrane O. A system for educational and vocational guidance in Morocco: Chatbot E-Orientation. In: FNC/MobiSPC; 2020. .

[5] Zuyeva TV, Nyssanov AT. Career guidance of adolescents in their sociocultural development and modern technologies. Psychologie Francaise. 2021.

[6] Pecht MG, Kang M. In: Machine Learning: Fundamentals; 2019. p. 85-109.

[7] Kashyap P. Machine Learning Algorithms and Their Relationship with Modern Technologies; 2017. .

[8] Ayodele TO. Types of machine learning algorithms. New advances in machine learning. 2010;3:19-48.

[9] Runkler T. Data Analytics: Models and Algorithms for Intelligent Data Analysis; 2020. .

[10] Zahour O, Benlahmar E, Eddaoui A, Hourrane O. A Comparative Study of Machine Learning Methods for Automatic Classification of Academic and Vocational Guidance Questions. Int J Interact Mob Technol. 2020;14:43-60.

[11] El Mrabet H, Moussa AA. IoT-school guidance: A holistic approach to vocational self-awareness & career path. Education and Information Technologies. 2021:1-18.

[12] El Mrabet H, Moussa AA. Smart School Guidance and Vocational Guidance System Through the Internet of Things. In: Proceedings of the 2nd International Conference on Networking, Information Systems and Security. NISS19. New York, NY, USA: Association for Computing Machinery; 2019. .

[13] Vignesh S, Shivani Priyanka C, Shree Manju H, Mythili K. An Intelligent Career Guidance System using Machine Learning. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). vol. 1; 2021. p. 987-90.

[14] Zahour O, Benlahmar E, Eddaoui A, Hourrane O. Automatic Classification of Academic and Vocational Guidance Questions using Multiclass Neural Network. International Journal of Advanced Computer Science and Applications. 2019;10.

[15] Jiménez-Raygoza LI, Medina-Vázquez AS, Pérez-Torres G. Proposal of a Computer System for Vocational Guidance with Data Mining. 2019 IEEE International Conference on Engineering Veracruz (ICEV). 2019;I:1-5.

[16] Kamburugamuwa P, Gamagae W, Dissanayaka W, Ahamd M, Abesiri L, Premadasa N. Smart Learning Guidance System for University Students. International journal of Enginerring Research and Technology(IJERT). 2020;9:689-94.

[17] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Commun ACM. 2017 May;60(6):84–90.