

A DROPOUT PREDICTOR SYSTEM IN MOOCs BASED ON NEURAL NETWORKS

Submitted: 26th June 2019; accepted: 25th March 2020

Khaoula Mrhar, Otmane Douimi, Mounia Abik

DOI: 10.14313/JAMRIS/4-2020/48

Abstract: *Massive open online courses, MOOCs, are a recent phenomenon that has achieved a tremendous media attention in the online education world. Certainly, the MOOCs have brought interest among the learners (given the number of enrolled learners in these courses). Nevertheless, the rate of dropout in MOOCs is very important. Indeed, a limited number of the enrolled learners complete their courses. The high dropout rate in MOOCs is perceived by the educator's community as one of the most important problems. It's related to diverse aspects, such as the motivation of the learners, their expectations and the lack of social interactions. However, to solve this problem, it is necessary to predict the likelihood of dropout in order to propose an appropriate intervention for learners at-risk of dropping out their courses. In this paper, we present a dropout predictor model based on a neural network algorithm and sentiment analysis feature that used the clickstream log and forum post data. Our model achieved an average AUC (Area under the curve) as high as 90% and the model with the feature of the learner's sentiments analysis attained average increase in AUC of 0.5%.*

Keywords: *Massive open online courses MOOCs, Student Attrition, Dropout prediction, Neural Network, Sentiment Analysis*

1. Introduction

More and more, the massive open online courses (MOOCs) have witnessed a tremendous development in the recent year [1]. Through the MOOCs, the learners have the opportunities to self-organize their participation, learning goals, knowledge, abilities and interests. Certainly, the MOOCs have brought interest among the learners (given the number of learners enrolled in these courses), nevertheless, There are many unresolved question related to MOOCs , One of the major recurring issues raised is the high dropout rate of the MOOC learners , Indeed, the MOOC dropout rate generally knows of 90 % attrition [2][3], in this sense, and according to the last study elaborate by EDX, only 17 % of the enrolled learners consulted the courses and only 8 % get the certification at the end of the MOOC [4], which means that most of the users who join a MOOC eventually don't complete it. However, the dropout in the MOOC was related to di-

verse aspects that can be classified as learner-related factors and MOOC-related factors. The learner-related factors are especially the motivation of the learner and their expectations [5][6][7], insufficient learner's background knowledge and lack of required learner's skills [8]. The MOOCs-related factors are related to courses design and the lack of the social interactions in MOOCs leading to an isolation feeling [9]. There has been an increasing attention to this issue by several researchers, in this sense, we can cite the paper of Kizilcec and al. [10], Hill [11] who proposed the classification of the learners according to their interactions with the platform. Hill [11] classified the various MOOCs participants in five categories: active, passive, Drop-ins, Observers and No. Shows. Besides, we can classify three different approaches to minimize the dropout rate [12][13]:

- The pedagogical strategy approach: it is a number of theoretical strategies validated by empirical studies. Among these educational strategies we can cite this work [14].
- The personalization and adaptation approach: several research works focus more and more on the importance of the personalization to reduce the dropout rate. In this context, we can cite different project which proposes a personalization system of the pedagogical objectives within MOOC [15] [16].
- The gamification approach: this is the mode of learning which capture a large audience. Indeed, integrating serious games into learning process in order to increase learners' motivation and engagement. [17].

However, to reduce the dropout rate, the adoption of these approaches is not sufficient. By reason of, the low pedagogical tutoring in these platforms, the no personalization of the courses according to the real learner profile (profile updated with the knowledge and skills acquired in MOOCs), thus the majority of the serious games couldn't influence the intrinsic and extrinsic learners' motivation. As consequence the need to predict the learner's dropout in MOOC in order to propose to them a suitable pedagogical strategy and/or complementary resources to help them to complete their courses. Our objective in this paper is to provide a solution to reduce the dropout rates through the prediction of the learner's dropout in the MOOC. Our approach consists to use a machine learning algorithm based on the clickstream log and

forum post data from EDX MOOC. We can summarize our research problems in three questions:

- How can we identify the features to be considered in the prediction task to have a better accuracy of our algorithm?
- What is the best machine learning algorithm to be used to predict the dropout of a learner?
- Can the sentiment analysis influence the accuracy of our algorithm?

2. Related Work

Recently, there have been several effort to predict learner dropout in MOOCs by analyzing the learner interaction, extracting a variety features and applying the machine learning algorithms. The mostly works use clickstream features which contain the interaction event among learners and the MOOCs courseware including discussion forum content, video lectures, quiz answers and more. Different from other works, using features such as number of threads viewed, number of forum posts, the number of video viewed and more to predict learner attrition [19].

Using learner's social interaction, quiz score and number of peer review [20]. And some works try to classify the various features to understand their relationship and their relative importance in order to classify the learners in MOOCs [21]. All these works use a variety of classification algorithms and adapt a different approaches for extracting features. In [22], the author propose a support vector machine SVM and extract features from clickstream to predict dropout learner each week. [23] Apply a logistic regression to identify the learners who seem to be not able to complete the course [23]. Author in [24] use a k-means to discover inactive learners in MOOCs environment. However, among various ML algorithms, only some researchers focus on Artificial Neural Network ANN [25] and Recurrent Neural Network RNN [26]. Table 1 presents a Synopsis of prior works on the dropout prediction in MOOCs.

Tab. 1. Survey of prior works on the dropout prediction in MOOCs

Study	MOOC Numbers	Dataset	Algorithm used
Balakrishnan et Coetzee [27]	1	Clickstream	HMM + SVM
Boyer et Veeramachaneni [28]	3	Clickstream	TL+LR
Chaplot et al. [25]	1	Clickstream forum, posts	ANN
Taylor et al. [29]	1	Clickstream, forum posts	LR
Coleman et al [30]	1	Clickstream	LDA+LR

Study	MOOC Numbers	Dataset	Algorithm used
Jiang et al.[31]	1	Socialnetwork, grades	LR
Kloft et al.[32]	1	Clickstream	SVM
Xing et al. [33]	1	Clickstream	PCA + {BN, DT}
Wang et al. [26]	1	Clickstream	CNN+RNN

However, several challenges facing dropout prediction using machine learning methods [34] (Fig. 1), such as the large mass of unstructured data contained in MOOCs platforms which need a specific management mainly when the missing data occurs. In this sense, we cannot apply several machine learning techniques that require a finite set of data and no missing observations such as HMM.

The non-interoperability of learning platforms and the non-standardization of data MOOC can lead to another challenge which the generalization of machine learning solution. Furthermore, each learning platform has its own data definition, vocabulary, and the clickstreams data are represented in a different format. Therefore, the process of creating, training and validation of ML model is specified of each learning platforms (Open edx, Coursera, canvas...) and can't be used for other learning platform.

Another challenge is the data variance in MOOC platforms that lead to imbalanced classes. The high data imbalance may result to poor performance, less accuracy and reliability in ML models such as SVM. In addition the privacy and non-availability of learner's data in MOOCs platforms and the availability only the clickstream data is one of the important challenge of ML learner dropout prediction in MOOCs because the result obtained limited and not representative

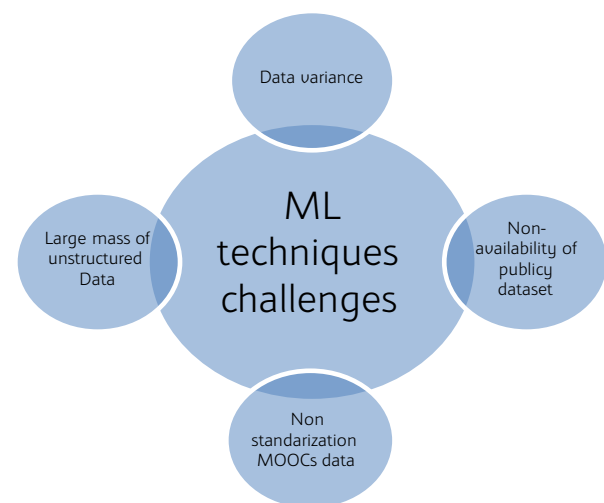


Fig 1. ML challenges for dropout prediction

Furthermore, according to some works [18], we observe a significant correlation between sentiment

expressed in the discussion forum of the course and the number of learners who drop the course. But unfortunately, there has not been much work on use of learner sentiments in predicting dropout. We propose in this paper a model that predicts learner attrition. This model is based on the most interesting information that impact the learner 'dropout' [34]. It's selected after a correlation with several features and learner 'dropout'. The feature is related to the clickstream log such as (Number of video views, number of subsection viewed ...) and related to discussion forum such as number of forum post viewed, number of forum post votes and student sentiments in discussion forum posts etc. However, we decided to use ANN model with two hidden layers and nine features input after several experimentation and accuracy comparison between the standard dropout prediction architecture. In addition, the choice of this algorithm is also to answer to some challenge of machine learning model cited before (large mass unstructured Data, solving high imbalanced data). In the following section, we introduce our proposed model and the machine learning algorithm used to address the MOOC dropout prediction problem.

3. Proposed Model

3.1. Prediction Dropout Problem Formulation

Since their appearance, the MOOC has several limits mainly the very high dropout rate, on all the MOOC, no matter their subject. On average, just 8% of the enrolled finish the courses and get a certification. The high dropout in MOOCs has been attribute to many factors, such as the lack of time, learner's motivation, feeling of isolation, and lack of interactivity in MOOCs. However, this problem raised the following research questions: "how can we predict the attrition of the learners in the MOOC?". To answer to this question, we propose a dropout predictor that can be used by the educators during the courses to propose necessary intervention for the learners at risk of dropping in order to reduce attrition learners in MOOCs we suggest applying four types of machine learning algorithm to the tracking log, discussion forum interaction clickstream and forum post data from OpenEdX Platforms MOOC 'Introduction to Computer Science', prepared exclusively for this work by Stanford's university. The objective of this step is to choose the algorithm which give the best accuracy in our context. We suggest afterward testing our algorithm by adding new features to improve even more the performance of the chosen algorithm. We also study in this paper, the influence of the sentiments analysis feature on the prediction accuracy

3.2 Dataset

The experimentation and analyses in this paper are based on a dataset which was prepared by Stanford's university exclusively for our work after a request via their official website (<https://data-stage.stanford.edu/>). The data was collected from a MOOC 'introduction to computer science' which was launched in March, 2016. The course lasted twelve weeks with 11 607 participants at the beginning of the week and 3861 participants staying until the last week of course. Globally, 20 828 learners are participated, with approximately ML techniques challenges Data variance No availability of public dataset Non interoperability of Learning Platforms Large mass of unstructured Data 81,4 % as dropout rate. Fig. 2 summarizes the various data sources received from Stanford:

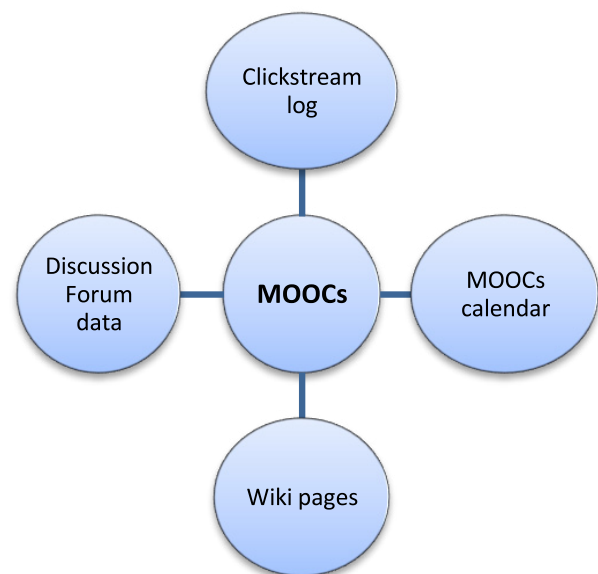


Fig. 2. Overview our Data sources

- A file of the clickstream log data from the historic navigation and edX server in CSV format. For example, every page visited by every learner was stored as event CSV.
- Forum post, comments and the answers stored in a file CSV.
- Wiki pages visited stored in a file CSV.
- A file CSV containing information about the state of the learners. For example, the database contained its final answer to a problem.
- A file CSV of the calendar of the MOOC which included information such as the deadlines of sending the homework.

We visualized several properties of our dataset in Fig. 3.

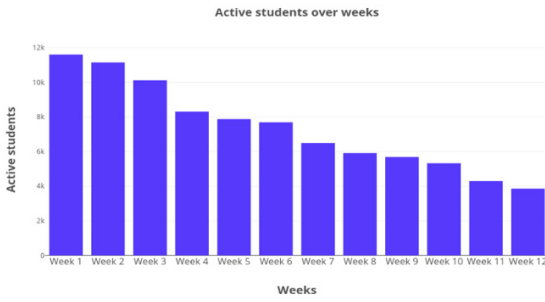


Fig. 3. Proprieties of our dataset

We observe that the number of active learners quickly decreases. Furthermore, the probability of dropout is high during the first two weeks.

3.3 Features Extraction

After the cleaning of our dataset, we process the data and transform them in the good format for the classification task. firstly, we translate the interactions in our data set in an adapted format to the task of classification. In particular, we analyze the interactions and extract a set of features, which are the input of our machine learning algorithms. Every features represents an aspect of the dataset which we want to take into account when we predict if a learner will dropout or not. To build our set of features, firstly, we sort out the interactions in the chronological order according to their timestamp, then, the missing attributes are replaced by the median of the week. After that, we select the features based on a manual feature selection method, we calculate a correlation between each pair of features and we choose nine features based on correlation matrix plot as appeared in figure 4.

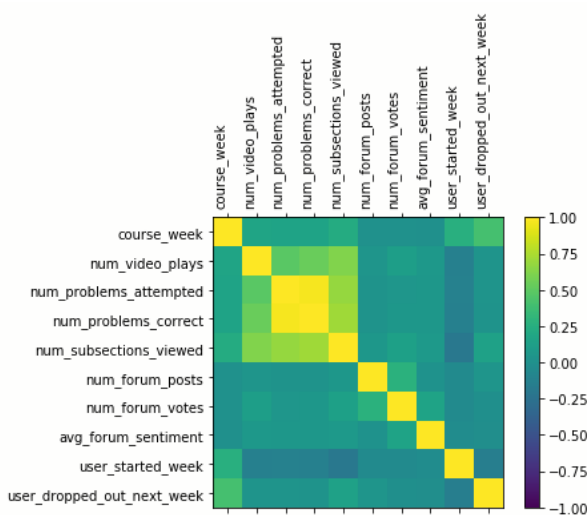
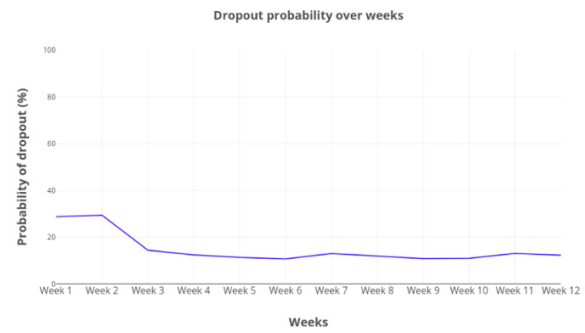


Fig. 4. Correlation Matrix Plot

The complete list of the extracted features is presented in Tab. 2.

After the extraction of these features, different important steps of preprocessing have been applied. Firstly, we applied several technique including the



removal of duplicate rows in order to cleaning our dataset. After that, the normalization procedure. Furthermore, the class imbalance is one of the issue that occurs in the MOOC dataset. Where the number of learners who dropout far exceeds the number of learners who complete the courses. The presence of the class imbalance in our dataset can bring the classifier to predict in an incorrectly way the learners may dropout. This is partially dangerous for the learners who are in the reality complete their courses and classified badly by our predicted model. To solve this problem under sampling the majority class ('learner dropping out ') was used.

Tab. 2. Features extracted for MOOCs dropout prediction

Features	Description
Course Week	The number of the weeks since the course has begun
Number of video views	The number of videos played by the learner in the current week
Number of problems answered	The number of the questions that the learner answered.
Number of problems answered correctly	The number of the questions that the learner answered correctly.
Number of subsection viewed	The number of section viewed by the learner in the current week
Number of forum post	The Number of forum pages viewed by the learner in the current week
Number of forum post votes	The Number of forum post voted by the learner in the current week
Learner Sentiment Analysis	The sentiment analysis of the learner in the forum post in the current week
Student started week	The number of the weeks since the learner began the course.

3.4. Proposed Model

To predict the learner attrition in MOOC, we suggest using the machine learning algorithm based on nine features presented in the previous section namely course week, Number of video views, Number of problems answered, Number of problems answered correctly, Number of subsection viewed , Number of forum

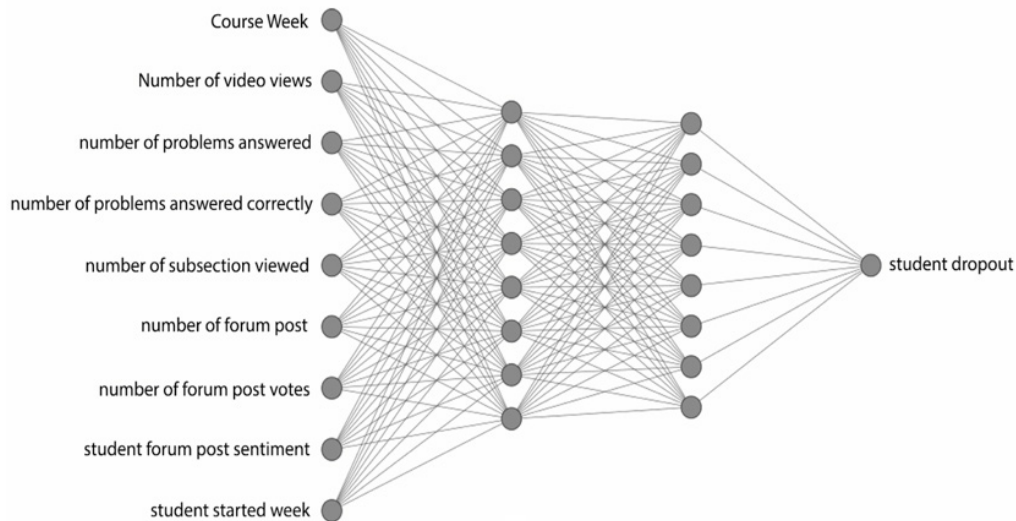


Fig. 5. Structure of our neural network used in the learner attrition

post, Number of forum post votes, Learner Sentiment Analysis, Student started week. So after the features extraction, we try to find the answers to these remaining questions:

- What is the best algorithm of machine learning to be used to predict the learner' dropout?
- The sentiment analysis feature can influence the accuracy of our learning algorithm?

In our work, we select several features related to discussion forum and clickstream based on a correlation between them and learners 'dropout. We decided to build an artificial neural network because is the most suitable to model the learner attrition prediction problem and there are a large number of inputs and any mathematical relation between the inputs and output is unknown contrary to many other machine learning techniques.

Furthermore, our neural network model consisting of nine nodes in the input layer which the value of the features for the current week. And the output layer is composed of one node which allow to predict if the learner will dropout the next week. Every input is normalized to take values between 0 and 1. We add two hidden layers with eight neurons in our neural network between the input and the output layers. The number of neurons in the hidden layer was determined experimentally to obtain the possible better results. Besides, to evaluate the prediction performance of the proposed model, we compare it with the others baseline algorithms used in MOOCs dropout prediction such as SVM, KNN, decision tree. Fig. 5 shows the structure of our neural network to predict the learner attrition.

To build the neural network, we use a resilient back propagation .it gave a better result in our experimentation in comparison with the back propagation and quick propagation. According to our experimentation concerning the four machine learning algorithms [table III], we notice that the neural network gives more accuracy comparing

with the others namely KNN, SVM and decision tree. So, we choose to use the neural network algorithm for our predictor dropout system. To respond to our research question concerning the impact of the sentiments analysis features on the accuracy of the results, we added the feature of the learner sentiments analysis in the discussion forum that already prepared in our dataset. Table 3 below summarizes the results of our experimentation of the four algorithms on our dataset.

Tab. 3. Comparison of AUC Average between learning algorithms

Algorithm	AUC Average
KNN	0.72
SVM	0.84
Decision Tree	0.74
Neural Network (Without Sentiments Analysis)	0.90
Neural Network (With Sentiment Analysis)	0.95

According to these results which will be discussed in details in future section results, we notice that the neural network algorithm has a better accuracy, besides, we noticed that the model considering the features "sentiment analysis" attained average increase in AUC of 0.05. This comparative study allowed us to consider that the neural network algorithm the most performing in the learner prediction attrition in MOOC.

4. Results

To measure the performance of our prediction algorithm, we use both metrics the AUC and the accuracy. AUC allow us to determine the validity of the

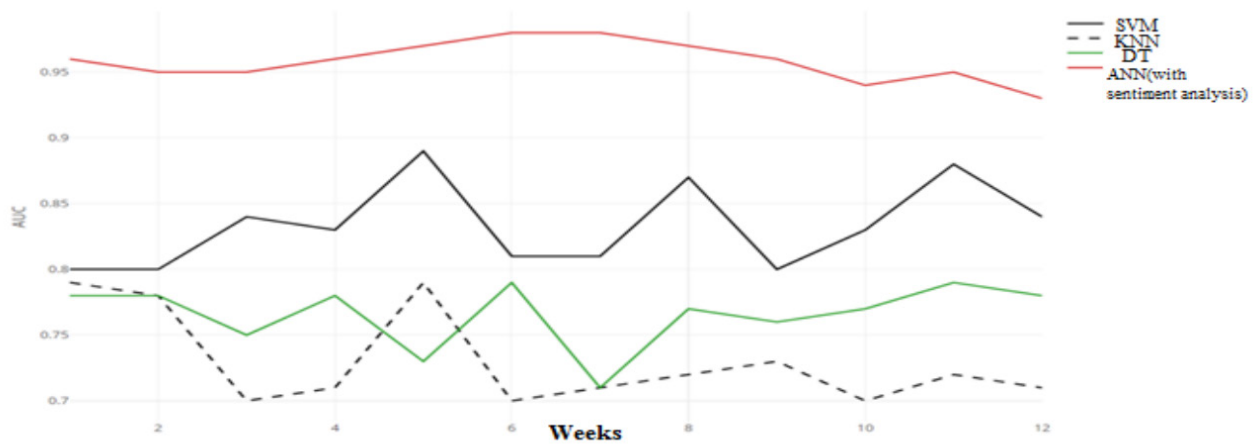


Fig. 6. Prediction performance for training data for AUC

Tab. 4. AUC and accuracy results for our algorithms overs weeks

Algorithm	Metric	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12
KNN	AUC	0.79	0.78	0.70	0.71	0.70	0.71	0.79	0.70	0.71	0.72	0.73	0.70
	Accuracy	0.82	0.83	0.87	0.90	0.89	0.86	0.86	0.82	0.82	0.83	0.80	0.81
SVM	AUC	0.80	0.80	0.84	0.83	0.89	0.81	0.87	0.80	0.83	0.89	0.88	0.84
	Accuracy	0.90	0.90	0.91	0.91	0.94	0.96	0.94	0.91	0.90	0.93	0.94	0.92
DT	AUC	0.78	0.75	0.78	0.73	0.79	0.71	0.77	0.76	0.77	0.79	0.78	0.79
	Accuracy	0.92	0.91	0.92	0.91	0.91	0.94	0.93	0.91	0.92	0.93	0.92	0.90
ANN (with the sentiment analysis feature)	AUC	0.96	0.95	0.95	0.96	0.97	0.98	0.98	0.97	0.96	0.94	0.95	0.93
	Accuracy	0.94	0.94	0.95	0.97	0.96	0.95	0.98	0.96	0.97	0.95	0.97	0.96

model, it give us the information about the probability of the test result, when the test is perfectly discriminating the surface under the curve (AUC) is 1.

This means that, for two learners (at risk learner of dropping and not at risk learner of dropping), the test allows us to distinguish between the learner who will dropout the MOOC and who will complete it. The table IV and the figure 6 present the performance of prediction of the various algorithms which we applied. The AUC value for KNN lies between 0.70 to 0.79 and for the decision tree, lies between 0.71 to 0.79. For SVM, AUC varied between 0.80 to 0.89.

For our algorithm which bases on the neural networks, the AUC varies between 0.93 to 0.98. The results of the AUC values show that the limit of prediction performance of the lowest prediction of our neural network has still exceeded the limit superior of the algorithms KNN and the decision tree. Our neural network has the highest average of the AUC values of 0.95.

It surpasses the baseline algorithms. In fact, in this study, most of the time a single layer was used in the construction of the machine Learning algorithm. There is enough space to refine the performances of our neural network.

Overall, our network of neurons presented a much better and performance for identification of the learners at-risk of dropping in MOOC.

Concerning the accuracy of the dropout prediction, the range for KNN was varied between 0.82 and 0.90 according to table 4. The range of precision for SVM was between 0.90 and 0.94. The interval for the decision tree was 0.90 and 0.94. For the precision of prediction of our algorithm, the range was 0.94 and 0.98. The precision of the neural network of neurons always has a better performance than the basic algorithms. According to this implementation we notice that the SVM has an accuracy prediction very close to neural network model.

5. Conclusion

The MOOC become more and more popular because of several elements that distinguish it to other online education courses (completely opened and free to anyone). However, a limited number of the enrolled learners complete the courses. Moreover, the MOOC teachers are incapable to identify the learners

at risk of dropping using the traditional methods (interview, observations and questionnaire).

Furthermore, the researchers began to explore the use of machine learning to develop a prediction dropout models for an effective identification of the learners at risk of attrition, and offer to them the suitable interventions. Nevertheless, in education context and in particular in MOOC, the data generated by the learners is enormous and varied. And the selection of the most informative features is the important task in the machine learning algorithms. In our work we choose nine features that are the most correlated with learner 'dropout related to discussion forum and clickstream data. And we construct an artificial neural network ANN model to predict if the learner will dropout the next week. According to our study, we conclude that in our context, the neural network is more performant than other baseline algorithm, including KNN, SVM and decision tree. By comparing the prediction performance of the various algorithms for the training data and test data, the results show that the prediction performance in neural network is more stable than the others. Besides, we note that the learner sentiment analysis in discussion forum feature improved the AUC value of our algorithm of 0.05. In future work, we plan to: Explore the social aspect of discussion forums in the detection of learner at risk of dropping. Improve more the performance prediction of our neural network by adding other features and hidden layers. Exploit the result obtained from our model to design a recommendation system which can help and motivate the learners to complete their MOOC.

AUTHORS

Khaoula Mrhar* – Faculty of Science, Mohammed V University in Rabat, Morocco, e-mail: khaoula_mrhar@um5.ac.ma.

Otmame Douimi – ENSIAS, Mohammed V University in Rabat, Morocco.

Mounia Abik – Information Retrieval and Data Analytics Research Team, ENSIAS, Mohammed V University in Rabat, Morocco.

*Corresponding author

REFERENCES

- [1] A. Watters, "MOOC Mania: Debunking the hype around massive open online courses", 2013, <http://www.thedigitalshift.com/2013/04/featured/got-mooc-massive-open-online-courses-are-poised-to-change-the-face-of-education/>. Accessed on: 2021-02-06.
- [2] K. Jordan, "Initial trends in enrolment and completion of massive open online courses", *The International Review of Research in Open and Distributed Learning*, vol. 15, no. 1, 2014, DOI: 10.19173/irrodl.v15i1.1651.
- [3] R. Meyer, "What It's Like to Teach a MOOC (and What the Heck's a MOOC?)", 2012, <https://www.theatlantic.com/technology/archive/2012/07/what-its-like-to-teach-a-mooc-and-what-the-hecks-a-mooc/260000/>. Accessed on: 2021-02-06.
- [4] D. F. O. Onah, J. Sinclair and R. Boyatt, "Dropout rates of massive open online courses: behavioural patterns". In: L. Gómez Chova, A. López Martínez and I. Candel Torres (eds.), *EDU-LEARN14 Proceedings*, 2014, 5825–5834.
- [5] Y. Belanger and J. Thornton, "Bioelectricity: A Quantitative Approach. Duke University's First MOOC", 2013, <https://dukespace.lib.duke.edu/dspace/handle/10161/6216>. Accessed on: 2021-02-06.
- [6] C. Gütl, R. H. Rizzardini, V. Chang and M. Morales, "Attrition in MOOC: Lessons Learned from Drop-Out Students". In: L. Uden, J. Sinclair, Y.-H. Tao and D. Liberona (eds.), *Learning Technology for Education in Cloud. MOOC and Big Data*, 2014, 37–48, DOI: 10.1007/978-3-319-10671-7_4.
- [7] P. Hill, "Emerging Student Patterns in MOOCs: A (Revised) Graphical View", 2013, <https://eliterate.us/emerging-student-patterns-in-moocs-a-revised-graphical-view/>. Accessed on: 2021-02-06.
- [8] H. Khalil and M. Ebner, "MOOCs Completion Rates and Possible Methods to Improve Retention - A Literature Review". In: *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications*, 2014, 1236–1244.
- [9] H. Khalil and M. Ebner, "'How satisfied are you with your MOOC?' - A Research Study on Interaction in Huge Online Courses". In: J. Herrington, A. Couros & V. Irvine (Eds.), *Proceedings of EdMedia + Innovate Learning*, 2013, 830–839.
- [10] R. F. Kizilcec, C. Piech and E. Schneider, "Deconstructing disengagement: analyzing learner subpopulations in massive open online courses". In: *Proceedings of the third international conference on learning analytics and knowledge*, 2013, 170–179.
- [11] P. Hill, "Some validation of MOOC student patterns graphic", 2013, <https://eliterate.us/validation-mooc-student-patterns-graphic/>. Accessed on: 2021-02-06.
- [12] A. Bakki, L. Oubahssi, C. Cherkaoui and S. George, "cMOOC: How to Assist Teachers in Integrating Motivational Aspects in Pedagogical Scenarios?". In: T. Brinda, N. Mavengere, I. Haukijärvi, C. Lewin and D. Passey (eds.), *Stakeholders and Information Technology in Education*, 2016, 72–81, DOI: 10.1007/978-3-319-54687-2_7.

- [13] A. Bakki, L. Oubahssi, C. Cherkaoui and S. George, "Motivation and Engagement in MOOCs: How to Increase Learning Motivation by Adapting Pedagogical Scenarios?". In: G. Conole, T. Klobučar, C. Rensing, J. Konert and E. Lavoué (eds.), *Design for Teaching and Learning in a Networked World*, 2015, 556–559, DOI: 10.1007/978-3-319-24258-3_58.
- [14] J. J. Williams, "Improving learning in MOOCs with Cognitive Science". In: *International Conference on Artificial Intelligence in Education 2013 Workshops Proceedings*, 2013.
- [15] S. Downes, "Creating the Connectivist Course", 2012, <https://halfanhour.blogspot.com/2012/01/creating-connectivist-course.html>. Accessed on: 2021-02-06.
- [16] F. Brouns, J. Mota, L. Morgado, D. Jansen, S. Fano, A. Silva and A. Teixeira, "A Networked Learning Framework for Effective MOOC Design: The ECO Project Approach". In: *Doing Things Better – Doing Better Things – EDENRW8 Conference Proceedings*, 2014, 161–172.
- [17] S. Deterding, D. Dixon, R. Khaled and L. Nacke, "From game design elements to gamefulness: defining "gamification"". In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 2011, 9–15, DOI: 10.1145/2181037.2181040.
- [18] M. Wen, D. Yang and C. P. Rosé, "Sentiment Analysis in MOOC Discussion Forums: What does it tell us?". In: *Proceedings of the 7th International Conference on Educational Data Mining*, 2014, 130–137.
- [19] T. Sinha, N. Li, P. Jermann and P. Dillenbourg, "Capturing "attrition intensifying" structural traits from didactic interaction sequences of MOOC learners". In: *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 2014, 42–49, DOI: 10.3115/v1/W14-4108.
- [20] M. Vitiello, S. Walk, R. Rizzardini, D. Helic and C. Gütl, "Classifying students to improve MOOC dropout rates", *Proceedings of the European Stakeholder Summit on experiences and best practices in and around MOOCs (EMOOCs 2016)*, 2016, 501–508.
- [21] S. Crossley, L. Paquette, M. Dascalu, D. S. McNamara and R. S. Baker, "Combining click-stream data with NLP tools to better understand MOOC completion". In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 2016, 6–14, DOI: 10.1145/2883851.2883931.
- [22] M. Kloft, F. Stiehler, Z. Zheng and N. Pinkwart, "Predicting MOOC Dropout over Weeks Using Machine Learning Methods". In: *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 2014, 60–65, DOI: 10.3115/v1/W14-4111.
- [23] J. He, J. Bailey, B. I. P. Rubinstein and R. Zhang, "Identifying at-risk students in massive open online courses". In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, 1749–1755.
- [24] T.-Y. Liu and X. Li, "Finding out Reasons for Low Completion in MOOC Environment: An Explicable Approach Using Hybrid Data Mining Methods", *Proceedings of 2017 International Conference on Modern Education and Information Technology (MEIT 2017)*, 2017, DOI: 10.12783/dtssehs/meit2017/12893.
- [25] D. S. Chaplot, E. Rhim and J. Kim, "Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks". In: *Proceedings of the Workshops at the 17th International Conference on Artificial Intelligence in Education AIED 2015*, vol. 3, 2015, 7–12.
- [26] W. Wang, H. Yu and C. Miao, "Deep Model for Dropout Prediction in MOOCs". In: *Proceedings of the 2nd International Conference on Crowd Science and Engineering*, 2017, 26–32, DOI: 10.1145/3126973.3126990.
- [27] G. Balakrishnan, "Predicting Student Retention in Massive Open Online Courses using Hidden Markov Models", Technical Report, 2013, <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-109.html>. Accessed on: 2021-02-06.
- [28] S. Boyer and K. Veeramachaneni, "Transfer Learning for Predictive Models in Massive Open Online Courses". In: C. Conati, N. Heffernan, A. Mitrovic and M. F. Verdejo (eds.), *Artificial Intelligence in Education*, 2015, 54–63, DOI: 10.1007/978-3-319-19773-9_6.
- [29] C. Taylor, K. Veeramachaneni and U.-M. O'Reilly, "Likely to stop? Predicting Stopout in Massive Open Online Courses", *arXiv:1408.3382 [cs]*, 2014.
- [30] C. A. Coleman, D. T. Seaton and I. Chuang, "Probabilistic Use Cases: Discovering Behavioral Patterns for Predicting Certification". In: *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 2015, 141–148, DOI: 10.1145/2724660.2724662.
- [31] S. Jiang, A. Williams, K. Schenke, M. Warschauer and D. O'Dowd, "Predicting MOOC performance with Week 1 Behavior". In: *Proceedings of the 7th International Conference on Educational Data Mining*, 2014, 273–275.
- [32] M. Kloft, F. Stiehler, Z. Zheng and N. Pinkwart, "Predicting MOOC Dropout over Weeks Using Machine Learning Methods". In: *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 2014, 60–65, DOI: 10.3115/v1/W14-4111.
- [33] W. Xing, X. Chen, J. Stein and M. Marcinkowski, "Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization", *Computers in Human Behavior*, vol. 58, 2016, 119–129, DOI: 10.1016/j.chb.2015.12.007.

- [34] F. Dalipi, A. S. Imran and Z. Kastrati, "MOOC dropout prediction using machine learning techniques: Review and research challenges". In: *2018 IEEE Global Engineering Education Conference (EDUCON)*, 2018, 1007-1014, DOI: 10.1109/EDUCON.2018.8363340.