Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

Optimizing Student Classification in Smart Education using Factor Analysis by Deep Boltzmann Machines and a Deep Learning Approach Improved by Bayesian optimization.

Abderrafik Laakel Hemdanou
AMCS team, Normal Higher School,
Abdelmalek Essaadi University
Morocco
abderrafik.laakelhemdanou@etu.uae.ac.ma

Mohammed LAMARTI SEFIAN
AMCS team, Normal Higher School,
Abdelmalek Essaadi University
Morocco
lamarti.mohammed.sefian@uae.ac.ma

Abstract— **Intelligent education aims to personalize learning according to students' individual needs. As student classification presents an essential task in this process, this research proposes an approach to optimize student classification using a combination of advanced techniques. Firstly, we use deep Boltzmann factor analysis to identify latent factors influencing student performance. These factors may include elements such as cognitive abilities, social characteristics, socio-economic characteristics and motivation levels. Secondly, we are developing a deep learning model enhanced by a combination of Bayesian and genetic methods. The Bayesian approach allows the incorporation of prior knowledge and the management of uncertainty and optimize the architecture and parameters of the deep learning model. The proposed approach is evaluated on a real student dataset and compared with traditional classification methods. The results show that our method outperforms existing approaches. In addition, latent factor analysis provides valuable insights into the key factors influencing student success. This research contributes to the advancement of intelligent education by providing a powerful tool for classifying students and personalizing learning.**

Keyword ;Optimization, Educationnal data mininig, Deep learning.

I. Introduction

The evolution of the education system has produced a large number of curricula and courses for students according to predefined standards. The student's journey begins with enrolment and ends with timely completion of the defined course. It's important that students take courses that match their interests, which will help them choose their future career. Academic advisors and students themselves want to know where they stand. Are they taking the right courses that will lead them to a successful career? The answers to these questions are obtained by classifying students. From the analysis of student classification[1], an advisor can compare the student's current level of achievement and future goal with the best estimate of the academic level they should achieve. Our overall goal is to optimize the student classification model. Optimization is necessary because a student always has no idea of his current position in the academic environment and his future position. We tried to determine the most important features using Deep Boltzman Machine[2], and then we optimized our DNN by defining a space of hyperparameters and optimizing them using Bayesian optimization.The research questions in this study can be summarized into three questions:

Q1: How does Factor Analysis contribute to student profiling using Deep Boltzman Machine ?

Q2: What advantages do Bayesian methods offer in deep learning applications?

Q3: How do Bayesian methods optimize deep learning models for student classification?

II. Factors importance

A. *Deep Boltzmann Machine*

Deep Boltzmann Machines (DBM) are unsupervised generative models used for learning complex representations of data. they are not directly designed for variable selection. However, an attempt was made to use an indirect approach to use DBMs in the context of variable selection. A DBM has been trained on the dataset which will allow the DBM to learn a representation of this dataset in its hidden layers[9]. and by analyzing connection weights such as connection weights between visible units (corresponding to input variables) and hidden units can indicate the relative importance of variables. whose variables with higher connection weights to hidden units could be considered more important. DBM's general response is given by:

$$E(v,h_k;\theta) = -\sum_{i,j} w_{ij}^{(v,h)} v_i h_i - \sum_i b_i^v v_i - \sum_j b_j^v h_j - \sum_{j,k} w_{jk}^{(hh)} h_j h_k$$

Where

- v represents the visible units.
- h represents the hidden units (with h_j and h_k representing different layers of hidden units)
- $w_{ij}^{(v,h)}$ is the weight between visible unit j and hidden unit k .
- $w_{jk}^{(hh)}$ is the weight between hidden unit j and hidden unit k in consecutive layers.
- b_i^v is the bias term for visible unit i.
- b_j^v is the bias term for hidden unit j.
- θ represents the parameters of the model, including the weights and biases.

The probability of a particular state is given by the Boltzmann distribution:

$$P(v,h;\theta) = \frac{e^{-E(v,h;\theta)}}{Z(\theta)}$$

Where $Z(\theta)$ is the partition function, a normalization factor that ensures the probabilities sum to one, and is defined as:

$$Z(\theta) = \sum_{v;h} e^{-E(v,h;\theta)}$$

In the context of variable selection, after training a DBM on a dataset, the connection weights between the visible units and the hidden units is analyzed to determine the importance of the input features

III. Optimization of DNN for classification

DNN Because of the significant amount of hidden layers[3], a deep neural network architecture is an artificial neural network architecture that is distinguished by its depth. The DNN can represent and capture intricate, non-linear relationships between data because to these intermediary layers. For a single neuron in a hidden layer, the fundamental formula of a deep neural network (DNN) looks like this:

$$a_i^{(l)} = \sigma\left(\sum_j (w_{ij}^{(l)} . a_j^{(l-1)}) + b_i^{(l)}\right)$$

where $a_i^{(l)}$ is the activation of the neuron i in the layer l, $w_{ij}^{(l)}$ is the weight associated with the connection between the neuron I in the layer l and the neuron j in the layer l-1, $a_j^{(l-1)}$ is the activation of the neuron j in the previous layer l-1, $b_i^{(l)}$ is the bias associated with the neuron i in the layer l, σ is the activation function.

to optimize the DNN, we defined a space of the hyperparameters to optimize of which there are two groups, the hyperparameters in relation to the architecture of DNN (Number of hidden layer, Number of neurons per layer, Layer type :dense, convolutional, recurrent,Activation function) and the others in relation to learning (Learning rate, Optimizer: Adam, SGD, RMSprop, Batch size, Number of epochs) [6].

A. *Bayesian optimization*

For Bayesian optimization, we chose The Tree-structured Parzen Estimator (TPE), TPE belongs to the family of Bayesian optimization methods and is designed to efficiently search the hyperparameter space by modeling the probability distribution of the hyperparameters given the observed performance[10]. TPE works by constructing two density functions $l(x)$ and $g(x)$, which correspond to the distribution of hyperparameters that result in "good" and "bad" performance, respectively. The performance threshold between "good" and "bad" is typically determined by a quantile of the observed performances. The algorithm then iteratively selects new hyperparameters to evaluate by optimizing the expected improvement (EI) over the current best performance, which is computed as follows:

$$EI(x) = \left(\alpha + \frac{g(x)}{l(x)} . (1-\alpha)\right)^{-1}$$

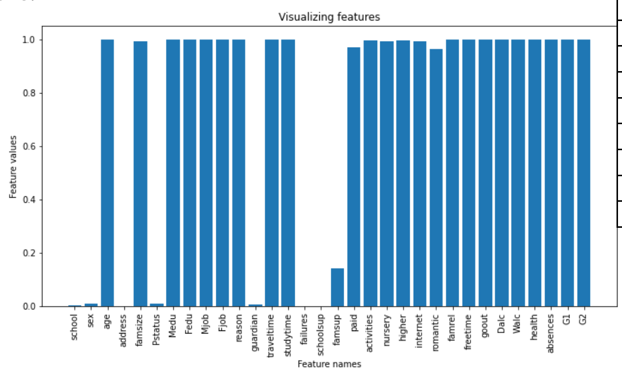
Where

- x represents the hyperparameters.

- α is the quantile that separates the "good" hyperparameters from the "bad" ones, based on the observed performances.
- $l(x)$ is the density function of the "good" hyperparameters.
- $g(x)$ is the density function of the "bad"hyperparameters.

IV. Result and discussion

The secondary education student results in two Portuguese schools are covered by the data used in this study. School reports and questionnaires were used to gather data on student grades as well as demographic, social, and school-related factors. Regarding performance in two different topics, mathematics and Portuguese language , two data sets are supplied. Here is a link to a description of each of the listed characteristics: https://archive.ics.uci.edu/dataset/320/student+performance. before building the DNN model optimized by Bayesian optimization[3], we tried to determine the importance of the most important features by adapting the DBM architecture to this task, as described in the section below. We obtained the results shown in the following figure.



the following most important factors: Age, famsize, Medu, Fedu, Mjob, Fjob, reason, traveltime, studytime, paid, activities, nursey, higher, internet, romantic, famrel, freetime, goout, Dalc, Walc, health, abscences, G1, G2.

This study revealed that factors linked to the family environment, such as parental education (Medu, Fedu) and parental occupation (Mjob, Fjob), are strongly correlated with academic achievement, suggesting that the socio-economic context in which students grow up can influence their academic performance. In addition, behavioral and lifestyle factors, such as studytime, extracurricular activities and (Dalc, Walc), also showed a significant influence, indicating that students' habits and lifestyle choices can have a direct impact on their academic performance.

Interestingly, factors such as access to the Internet (internet) and the aspiration to pursue higher education (higher) were associated with better performance, underlining the importance of educational resources and academic goals in students' school careers. In contrast, factors such as romantic relationships and going out (goout) appear to have a negative impact, which could reflect distraction or conflict between social commitments and academic responsibilities.

The DBM results offer valuable insights for education stakeholders, who can target these factors in the development of policies and programs aimed at improving academic success.

Classification models have been applied to data set and DBM subset, and the evaluation metrics used are defined as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad Precision = \frac{TP}{TP+FF}$$

$$Recall = \frac{TP}{TP+FN} \qquad \mathrm{F1-Score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Reacll}}$$

With TP: True Positives, TN: True Negatives, FP: False Positives, FN: False Negatives

The table below shows the different results obtained from each model on the data and on the subset of DBM .

TABLE I. RESULTS OF DIFFERENT MODELS

Model		**Student**	**Data**		
	Accuracy	**Precision**	**F1-Score**	**Reacll**	**Roc-Auc**
KNN	0.62	0.66	0.74	0.84	0.55
SVC	0.68	0.69	0.79	**0.94**	0.66
DNN	0.69	0.70	0.80	0.92	0.69
B DNN	**0.74**	**0.73**	**0.81**	0.92	**0.72**
		DBM	Subset		
KNN	0.55	0.62	0.70	0.80	0.52
SVC	0.63	0.64	0.77	**0.96**	0.51
DNN	0.62	0.65	0.75	0.88	0.59
B DNN	**0.70**	**0.71**	**0.79**	0.91	**0.65**

The following two figures show the ROC(Receiver Operating Characteristic) of the models to compare our model with other models.

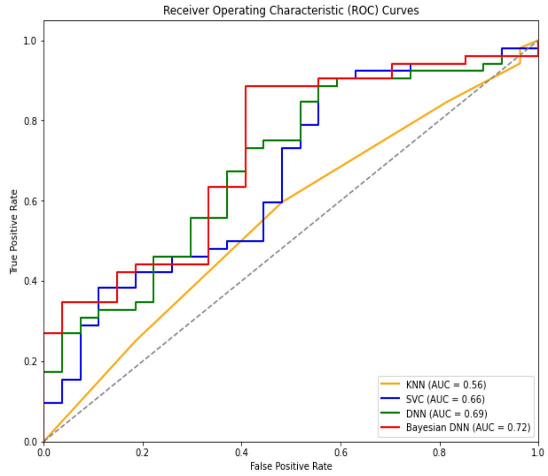


Figure 1. The ROC (Receiver Operating Characteristic) curve on student data

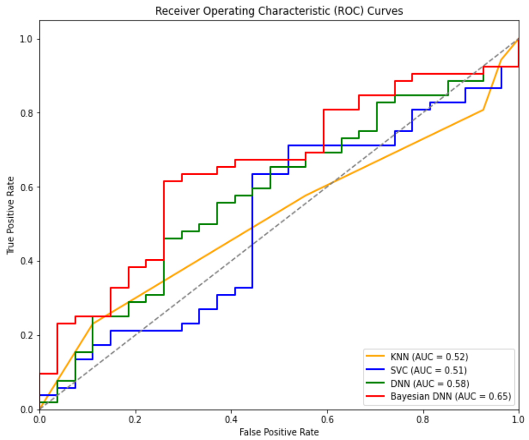


Figure 2. The ROC (Receiver Operating Characteristic) curve on subset DBM

The performance of four different models—KNN, SVC, DNN, and Bayesian DNN—was evaluated on two datasets, referred to as the Student Dataset and Subset DBM. The evaluation metrics included Accuracy, Precision, F1-Score, Recall, and Roc-Auc.
For the 'Student' data, the Bayesian DNN outperformed the other models across all metrics, achieving the highest Accuracy (0.74), Precision (0.72), F1-Score (0.81), Recall (0.92), and Roc-Auc (0.72). The DNN model followed closely, with a slightly lower performance, particularly in Accuracy (0.69) and Roc-Auc (0.69). The SVC model showed a moderate performance with an Accuracy of 0.68 and the lowest Roc-Auc of 0.66 among the top-performing models. The KNN model had the lowest scores in this dataset, with an Accuracy of 0.62 and Roc-Auc of 0.55.
In the 'Subset DBM' dataset, the Bayesian DNN again showed the best performance, with the highest Accuracy (0.70) and Roc-Auc (0.65). The DNN model's performance dropped compared to the 'Student' dataset, with a lower Accuracy of 0.62 and Roc-Auc of 0.59. The SVC model, while having a higher Accuracy (0.63) than the DNN, had the lowest Roc-Auc (0.51) in this dataset. The KNN model had the lowest Accuracy (0.55) and Roc-Auc (0.52) among the evaluated models.

The results indicate that the Bayesian DNN model is the most robust across both datasets, consistently achieving the highest scores in all metrics[4]. This suggests that the Bayesian approach to DNN may provide a more reliable prediction, potentially due to its ability to better manage uncertainty and prevent overfitting compared to traditional DNN[7].
The DNN model performed well on the Student Dataset but showed a decrease in performance on the Subset DBM Dataset. This could be due to differences in the complexity or size of the datasets, indicating that the DNN may require further tuning to maintain high performance across varying data conditions.
The SVC model demonstrated good Recall, especially in the 'Subset DBM' dataset, suggesting that it is particularly adept at identifying the positive class. However, its lower Roc-Auc scores imply that there may be a trade-off with its ability to correctly identify the negative class.
The KNN model had the lowest performance in both datasets, which might be due to its sensitivity to the choice of 'k' and the distance metric used.in general, the superior performance of Bayesian DNN could be attributed to its probabilistic nature, which enables a more nuanced understanding of model uncertainty. this opens up another avenue of research to exploit this optimization method for other machine and deep learning methods[5].

V. Conclusion

The findings of the study highlight the Bayesian DNN as the most effective model, consistently outperforming the others in terms of all the metrics considered. Its superior performance underscores the benefits of incorporating Bayesian inference into deep learning, particularly in handling uncertainty and providing more. While the DNN model showed promising results, particularly with the 'Student' dataset, its performance varied with the 'Subset DBM' dataset, suggesting a potential need for adaptive hyperparameter tuning when dealing with different types of data. The SVC model's strong Recall indicates its capability in identifying positive cases, but its lower Roc-Auc scores point to limitations in distinguishing between classes effectively.

The KNN model, although a simpler algorithm, performed less well, which can be attributed to its sensitivity to hyperparameters and the nature of the data sets[8]. This suggests that more complex models or feature engineering may be required to improve its predictive ability. [10]

In conclusion, the study demonstrates the importance of model feature selection in machine and deep learning tasks, with Bayesian DNN emerging as an accurate choice for handling diverse datasets. Future work could explore the integration of these models with other techniques, and optimize other models using Bayesian optimization, to further improve their performance and applicability to a wider range of data-driven problems.

References

[1] Mohammed Afzal Ahamed, Prayuk Chaisanit, and TR Mahesh. Performance of student prediction. International Journal of Computer Science and Information Technology, 8(6), 2019.

[2] Nicholas Robert Beckham, Limas Jaya Akeh, Giodio Nathanael Pratama Mitaart, and Jurike V Moniaga.Determining factors that affect student performance using various machine learning methods. Procedia Computer Science, 216:597−603, 2023.

[3] Mohammed El Fouki, Noura Aknin, and Kamal El Kadiri. Multidimensional approach based on deep learning to improve the prediction performance of dnn models. International Journal of Emerging Technologies in Learning, 14(2), 2019.

[4] Aya Nabil, Mohammed Seyam, and Ahmed Abou-Elfetouh. Prediction of students’ academic performance based on courses ’ grades using deep neural networks. IEEE Access, 9:140731−140746, 2021.

[5] Kingsley Okoye, Julius T Nganji, Jose Escamilla, and Samira Hosseini. Machine learning model (rg-dmml) and ensemble algorithm for prediction of students’ retention and graduation in education. Computers and Education: Artificial Intelligence, 6:100205, 2024.

[6] Kukkar A, Mohana R, Sharma A and Nayyar A. (2024). A novel methodology using RNN + LSTM + ML for predicting student's academic performance. Education and Information Technologies. 10.1007/s10639-023-12394-0.

[7] Rahul and Katarya R. (2024). A Systematic Review on Predicting the Performance of Students in Higher Education in Offline Mode Using Machine Learning Techniques. Wireless Personal Communications. 10.1007/s11277-023-10838-x. 133:3. (1643-1674). Online publication date: 1-Dec-2023.

[8] P. P and Veeramanju K. (2023). A Systematic Review on Machine Learning Algorithms for Customer Satisfaction Classification in Various Fields. International Journal of Management, Technology, and Social Sciences. 10.47992/IJMTS.2581.6012.0305. (326-339).

[9] Boltzmann stacked classification data mining model for the student performance improvement in academic performance, Sugin Lal, G and Porkodi, R,Journal of Intelligent & Fuzzy Systems, number={Preprint},(1—17), IOS Pres.

[10] Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for imbalanced tabular data,Tao, Shuting and Peng, Peng and Li, Yunfei and Sun, Haiyue and Li, Qi and Wang, Hongwei,Expert Systems with Applications}, Vol(237),(121-294),(2024), Elsevier.