EDRAAK
Vol. (2023), 2023, pp. 1–4
ISSN: xxxx-xxxx

Research Article

predictive analytics model for students' grade prediction using machine learning. A review

Muhammed Fareed Flayyih[1,*], Hassan TOUT[2],

[1] Computer Science Dep., Arts, Sciences and Technology University in Lebanon, Beirut, Leonean.

[2] Computer Science Dep., Lebanese University in Lebanon, Beirut, Leonean.

ARTICLE INFO

Article History
Received 25 Sep 2022
Revised: 20 Nov 2022
Accepted 23 Des 2022
Published 15 Jan 2023

Keywords
Academic achievement,
Machine learning,
Neural Networks,
Predictive accuracy,
Educational data.

ABSTRACT

Forecasting the academic achievement of students is a vital undertaking in the analysis of educational data, providing the opportunity to improve learning results and teaching methods. This paper investigates an array of machine learning (ML) methods for predicting student grades, offering a thorough examination of cutting-edge methodologies. The research centres on three primary projects that utilise supervised, unsupervised, and reinforcement learning techniques, namely Logistic Regression, Linear Discriminant Analysis, Random Forest, and Neural Networks. A comparative examination demonstrates that Neural Networks often surpass other models in terms of their predictive accuracy, adaptability, and capacity to represent intricate, nonlinear relationships in educational data. Notwithstanding issues like overfitting and the requirement for extensive training data, Neural Networks exhibit considerable promise for wider applications in personalized learning and academic prediction. The findings emphasize the profound influence of machine learning in educational environments, promoting the use of advanced algorithms to facilitate data-driven decision-making in academics.

1. INTRODUCTION

Within the ever-changing field of educational data analysis and academic forecasting, the goal of precisely predicting students' performance is of utmost importance. In this study, we will investigate the current state of the art by focusing on three specific initiatives that aim to forecast students' grades using machine learning algorithms. The analysis is organized into four separate sections, each highlighting a particular aspect of this complex problem[1].

The first section of the paper delves into several machine learning algorithms and activation functions, emphasizing their importance in training prediction. This study aims to enhance our comprehension of these algorithms and the significance of their use in the exemplary projects showcased in the second section. The latter part of the paper focuses on predictive analytics, specifically analyzing three initiatives that aim to predict student academic success. The examination highlights advanced algorithms and data-driven approaches[1].

This section presents a comparative examination of the machine learning models, specifically examining their advantages and disadvantages in forecasting student achievement. This section concludes by identifying the prevailing machine learning algorithm utilized in these studies and examining its wider implications for academic forecasting.2. Methods of applied machine learning[1].

2. MACHINE LEARNING TECHNIQUES

*Corresponding author email: mfa058@live.aul.edu.lb
DOI: https://doi.org/10.70470/EDRAAK/2023/001

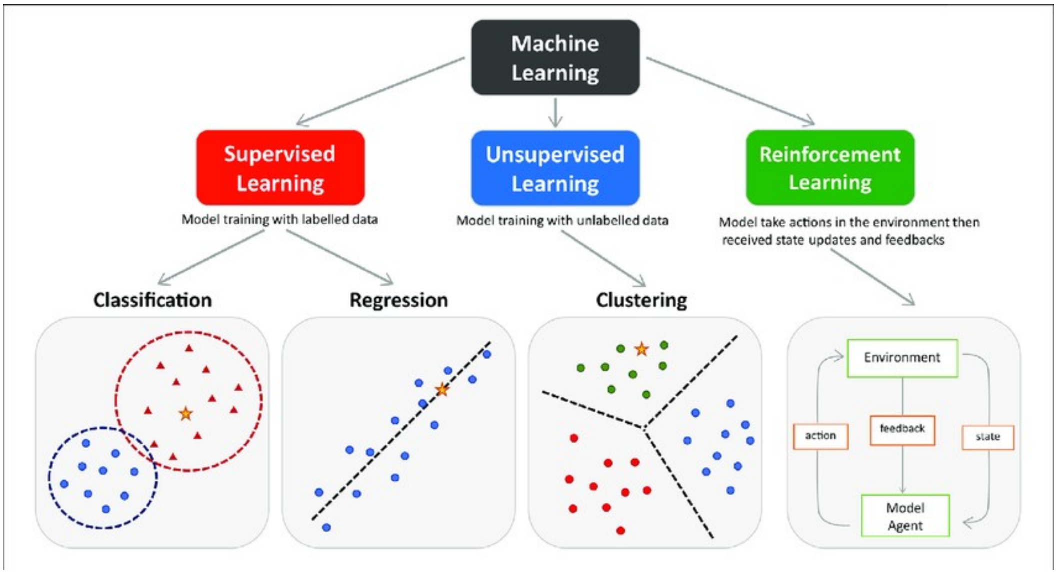

Fig .1. Inflation neural network diagram

1. Supervised learning:
   Investigation of algorithms that acquire knowledge from annotated data in order to generate predictions or classifications. Typical methodologies encompass[2]:
   Classification refers to the prediction of distinct class labels, such as spam detection or sentiment analysis.
   Regression is a statistical method used to estimate continuous data, such as home prices, by considering characteristics such as size and location[2].

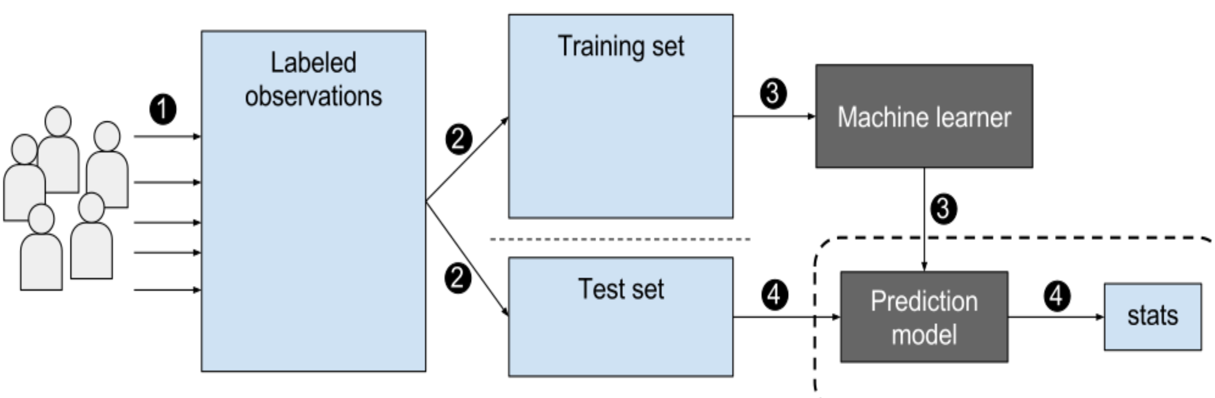

Fig .2. supervised learning

2. Unsupervised learning:
   The primary objective is to detect patterns in unlabeled data using methods such as:
3. Clustering: Categories comparable data points (e.g., K-Means).
4. dimensionality reduction: Minimizes the number of features while preserving crucial information (e.g., Principal Component Analysis and t-SNE)[3].

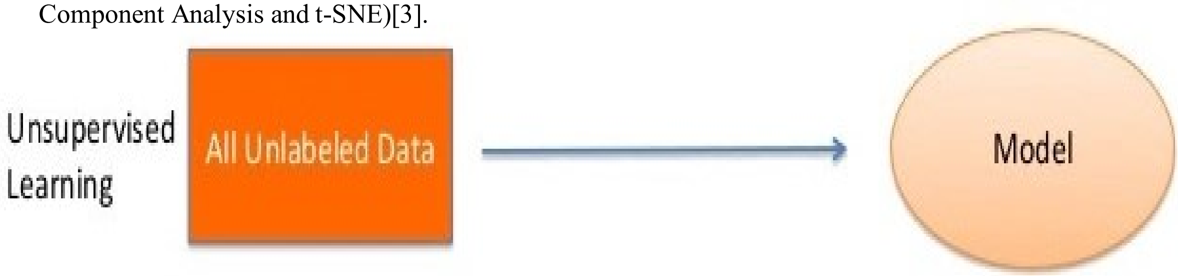

Fig .3. Unsupervised Learning

5. Reinforcement learning:
   This paper presents a technique in which an agent acquires optimal actions by means of trial and error. The paper emphasizes methods such as Q-Learning and Policy Gradient Methods, which are widely employed in the fields of robotics, gaming, and autonomous vehicles[3].

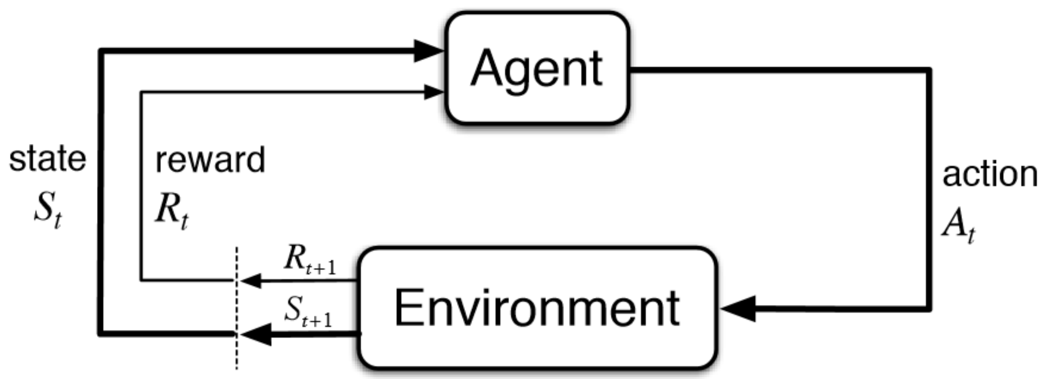

Fig .4. Reinforcement Learning

3. LAYERS OF NEURAL NETWORKS

The Perceptron Layer is the central component of neural networks, employing several activation functions such as ReLU, hyperbolic tangent, and logistic activation functions to represent intricate input patterns[4].
In classification tasks, the probabilistic layer is responsible for generating output probabilities that guarantee the total of all class probabilities is equal to one[4].
The Long-Short-Term Memory (LSTM) Layer processes sequential data, such as time series, by capturing long-term relationships via distinct gates (forget, input, state, and output)[5].
In order to improve learning and generalization, input data is normalized using scaling and unscaling layers. The purpose of the unscaling layer is to restore model outputs to their initial scales for subsequent interpretation[5].
The bounding layer is a component that limits the outputs of a neural network to predetermined ranges. It is particularly valuable in applications such as quality evaluations or financial forecasts[5].

4. COMPARATIVE EVALUATION OF STATISTICAL MODELS

This section provides a comparison of different machine learning models employed in three educational projects.
Logistic Regression is a statistical method used to classify student performance into separate unique categories (Distinction, Fail, Pass) by considering various input features[6].
Linear Discriminant Analysis (LDA) is a statistical method used to evaluate the efficacy of linear combinations of variables in differentiating between different classes[6].
The Random Forest algorithm incorporates ensemble learning to enhance accuracy by constructing many decision trees to forecast student performance[7].
Neural Networks demonstrate superior predictive accuracy, particularly in identifying intricate, non-linear patterns in student data. Nevertheless, it encountered difficulties in forecasting particular instances of failure, necessitating additional refinement[7].

5. ANALYSIS OF EXEMPLARY PROJECTS

Project 1: Applying Neural Networks, Logistic Regression, Latent Dirichlet Allocation (LDA), and Random Forest algorithms to predict academic results. Neural Networks performed exceptionally well, albeit with certain challenges in predicting failures[8].
Initiative 2: Utilizes TensorFlow and neural networks to forecast academic achievement in Portuguese and Math disciplines, by utilizing ReLU activation and dense layers to accurately represent student data[8].

Research Project 3: Centers on forecasting students' ultimate examination outcomes using K-Nearest Neighbors (K-NN), emphasizing its straightforwardness but also the computational difficulties arising from distance computations across extensive datasets [8].

6. CONCLUSION

Neural Networks are distinguished by their exceptional precision, flexibility, and capacity to represent complex features in data, rendering them very appropriate for forecasting student achievement.
Furthermore, they demonstrate potential in wider educational contexts, such as individualized learning, policy enhancement, and focused interventions derived from predictive analysis.

References

[1] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. University Book, 2021.
[2] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H. Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," Education and Information Technologies, vol. 28, no. 1, pp. 905–971, Jan. 2023.
[3] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 3, p. e1355, 2020.
[4] Y. Zhang and H. Rangwala, "Deep learning based multi-task learning for online student performance prediction," IEEE Transactions on Knowledge and Data Engineering, vol. 32, no. 2, pp. 283-293, 2018.
[5] H. Hussein and A. Hassan, "Predicting student performance using artificial neural networks: A case study in higher education," International Journal of Advanced Computer Science and Applications, vol. 12, no. 5, pp. 112-118, 2021.
[6] H. Yu, X. Wu, and W. He, "Comparative study of machine learning algorithms for student academic performance prediction," Journal of Educational Computing Research, vol. 57, no. 1, pp. 208-227, 2019.
[7] Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 2, pp. 26-31, 2018.
[8] D. Saxena, "Student performance prediction: Complete analysis," Kaggle, Available: https://www.kaggle.com/code/devassaxd/student-performance-prediction-complete-analysis.