



Reducing college dropout ratio using Machine Learning

• Rakesh Kumar • Anshuman

Received : April 2022

Accepted : May 2022

Corresponding Author : Rakesh Kumar

Abstract: Machine Learning (ML) has proved significant in almost all walks of life. The presented work focuses on the implementation of ML algorithms in the field of humanity and education. Every year a large number of students choose to drop their studies before completing college. Some of them build their own future as entrepreneurs and some struggle throughout their lives to meet the ends. In both cases, the institute or some deserving candidates suffer. This is becoming a serious concern, as all over the world the dropout rates are significantly high. This work focuses on the notion of implementation of Machine Learning algorithms to classify the traits of various students based on their activities and past performances to describe the dropout ratio and also certain measures to reduce the numbers.

Keywords: *Machine Learning, Artificial Intelligence, Supervised Learning, Classification, Clustering.*

Rakesh Kumar

Faculty – Computer Applications,

LNMI, Patna

Email-id: rakeshkr1@lnmipat.ac.in

Anshuman

Assistant Professor,

Amity University Patna

Email-id: aaryaa@ptn.amity.edu

Introduction:

Machine Learning (ML) as a sub-domain of Artificial Intelligence (AI) is doing wonders in many fields. There is an endless number of examples of implementation of ML algorithms and techniques in an innovative manner to a situation, that lead to more than desired results. Whether we talk about self-driving cars or Nanny robots, AI works like the eye of a guardian to take care of the belongings when it is most difficult for a human to do continuously the same boring task for a very long period of time. And not just in these scenarios, where monitoring is considered to be whole day task, but also for a longer period of time, such as the whole year or even a decade. Machine Learning allows us to dive deep into the data generated by humans and my precious information that would do wonders with respect to their health issues, careers, and whatnot.

Machine Learning has three basic implementations viz. supervised learning, unsupervised learning and reinforcement learning. We like to extend the notion of supervised learning in the present work. A supervised learning algorithm needs a set of data on which the algorithm is trained and then tested for accuracy. Usually, an 80-20 ratio meets standard benchmarks for such algorithms for training and then for the testing purpose respectively. Regression, Decision Tree, Random Forest, and Classification are the names of some popular supervised machine learning

techniques. The presented work talks about the notion of implementation of the classification algorithms for predicting the traits of a person as a student to define how much chance is there that such a person would drop out from studies at a certain level of difficulty. For this purpose, Support Vector Classifier (SVC), Logistic Regression Classifier (LRC), and Naïve Bayes Classifier (NBC) would serve perfectly. Each one of them has its own way of zeroing the results into one of the output classes, in this case, the answer to the question “Whether someone takes a dropout or not”.

Objective:

The work presented here aims at making a prediction about a person based on certain traits in the previous time span. For example – A college student in their graduation leaves the course before completion because of the increased complexity of the course topics or even due to long traveling time every day to a distant institute or maybe because of a not so up to the mark company during that time period. The idea is to develop an algorithm that would not only collect significant data about a person but further filters it for precise calculations and predictions during a critical event. It has the potential to work as an AI counsellor for students that tells what courses to opt for and what not, based on the interest and abilities of a student. The usual process requires some sort of questionnaire answered by the candidate, but here the algorithm takes care of the data like a personal assistant and then guides or advises as and when needed.

Methodology:

The work starts with the collection of data about an individual. That can be done via a smart personal assistant program or smart wearables like smart watches or smart bands or even smartphones allowed to capture minute data values. These values need to be stored safely for a longer period of time, even for a lifetime of a person. Cloud storage or backup facility would serve the purpose. All the data aggregated can then be used to generate important details based on various traits of that person, like what should they eat to stay happy if the situation around is not very favorable. Similarly, “What courses to opt for during the admission

to a certain degree program”[2]. That can be done after preprocessing the data and implementing an ML algorithm to make a strong prediction based on the data available from the worldwide cloud networks. The workflow notion is depicted in Fig. 1.



Fig. 1. Workflow of the model

The prediction so obtained be further used, to either provide timebound support to that person in case of some unfavorable situation or if handled more minutely may also avoid such situations. Like extra support/advice, while making an important decision, such predictions may be proven very lifesaving to some persons. Students even commit suicide because the pressure is not handled correctly [1]. The dropout is still the second concern. However, at the initial level, the model would generate data for probable dropouts based on early traits [5].

Future scope:

A lot can be done in this field, as almost all the companies accessing the personal data of their customers focuses to sell it out for making an easy profit, which in turn results in some sort of harassment of the person. Focusing rather to support and develop better qualities in a person in exchange for lesser profit is one thing or it could become such a revolution that better organizations or even governments play important roles in the implementation in the future. A person learning what they enjoy the most would implicitly reduce the dropout problems.

Conclusion:

The implementation of ML algorithms for a completely new sector needs very careful monitoring, as

it serves endless benefits on one hand, and on the other, it may lead to harassment if falls into wrong hands [4]. Collecting and using a person's data for advising them throughout their life for better decision making seems a very vague idea at first but the implementation is very interesting and promising. Further development would lead the path to such a society, where people would not judge each other on the basis of decisions imposed on themselves, but rather with healthy choices to develop healthy societies.

References :

1. Abdelhamid N., and Thabtah F. (2014). Associative Classification Approaches: Review and Comparison. *Journal of Information and Knowledge Management (JIKM)*, 13(3).
2. Cao, C. (2012). Sports data mining technology is used in basketball outcome prediction. Master's Thesis, Dublin Institute of Technology, Ireland.
3. Delen, D., Cogdell, D., & Kasap, N. (2012). A comparative analysis of data mining methods in predicting NCAA bowl outcomes. *International Journal of Forecasting*, 28(2), 543-552.
4. Nunes, S., and Sousa, N. (2006). Applying Data Mining Techniques to Football Data from European Championships. *Proceedings of Conference de Metodologias de Investigação Científica*, 4-16
5. Wiseman, O. (2016). Using Machine Learning to Predict the Winning Score of Professional Golf Events on the PGA Tour (Doctoral dissertation, Dublin, National College of Ireland).