International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V68I6P110

Volume 68 Issue 6, 57-63, May 2020

Original Article

Maintain and Evaluate Students Performance Using Machine Learning

Mujthaba G.M [1], Gulam Mubasheer Mujtaba[2], Mohammed Rahmath [3]

[1]*Lecturer in Computer Science Department, Prince Sattam Bin Abdul-Aziz University, K.S.A.*
[2] *Assistant Professor in Electronics and Communications Department, Methodist College ofEngineering & Technology, Osmania University, India.*
[3]*Lecturer in Computer Science Department, Prince Sattam Bin Abdul-Aziz University, K.S.A.*

Received Date: 03 May 2020
Revised Date: 14 June 2020
Accepted Date: 16 June 2020

Abstract *- Machine learning is an obligatory branch of Artificial Intelligence in computer science that is used to analyze data, evaluate analytical models, visualize & predict outputs and decisions making. These operations require the best learning approaches, which hassles machine learning algorithms. The numerous machine learning algorithms are categorized into primarily supervised, unsupervised and reinforcement algorithms. The main objective of these algorithms is to lessen human interactions with appropriate decision making. This paper highpoint how these machine algorithms can be beneficial in students learning. The effective use of supervised machine algorithms like regression, decision trees, and logistic representations will be applied to learning classrooms for higher education. To implement these algorithms, this research paper chooses the R programming language as R is known to be an insightful and influential language in creating data sets, identifying patterns, modelling & visualizing data, predicting and making proper decision making. These machine learning algorithms will transform the student's data into appropriate decisions. All the learning operations are performed by using the R language and generating the results.*

Keywords *- Machine Learning (M.L.), Linear Regression, Multiple Regression, Decision trees, Logistic representations and R Programming.*

I. INTRODUCTION

Artificial intelligence is a branch of computer science that is the so-called predominant technology that creates an interactive environment between its users through the internet. The overwhelming existing learning mechanism that comes through artificial intelligence is machine learning [2, 20]. A. I am the vast field that transforms user interactions from physical to logical by means of the digital world.

A. Importance of machine learning

The main objective of these is to use machine learning algorithms of A. I am to reduce the user's interactions, best learning techniques to solve the problem, precise outputs, incorporate statistical analytical approaches and grow organizations economically [2,3,4] as growing M.L provides reliable products & services to their users in various organizations to fulfil social impacts, transparency and responsibility towards its users. These M.L algorithms are extensively used in engineering products, retails, health & medical sciences, financial institutions, emergency services, travel & tourism, educational & research growth and developmental life sciences [2,20]. The effective use of data by the data scientists completely relies on how efficiently the M.L algorithms are devised & designed to meet the needs of their users. While deploying the M.L algorithms, users need to be kept in mind how the systems can act like human beings and use technology according to it. Advancements in Artificial intelligence demand and further classification into mainly supervised, unsupervised and reinforcement algorithms. Supervised algorithms are used to train the labelled data sets, which evaluates more accurate. At the same time, an unsupervised learning algorithm is to train the unlabeled data, which may not be more accurate. Another important learning is reinforcement learning which is dynamic in nature that trains the software agents how to behave or interact in the environment. The best examples for the supervised learning algorithms are Linear & multiple regression, logistic representations, decision trees which is well demonstrated in the below sections. Unsupervised learning algorithms are classified into K- means of clustering problems [21]. Apriori algorithms, principal component analysis and many more. At the same time, reinforcement algorithms are classified into Q-Learning lambda, SARSA lambda, DQN and others.

B. Machine learning model

As stated above, the M.L algorithms main perspective is to caricaturist the pattern of resignation. Pattern recognition deals with numerous activities like Pattern acquisitions, training the model with appropriate M.L algorithms, identifying new patterns and applying the models to various applications.

The following below figure 1 demonstrates the steps which are carried out by the M.L [2] [3] [4] and

[24] the below fig 1 illustrate the Machine Learning model and consists of the below steps

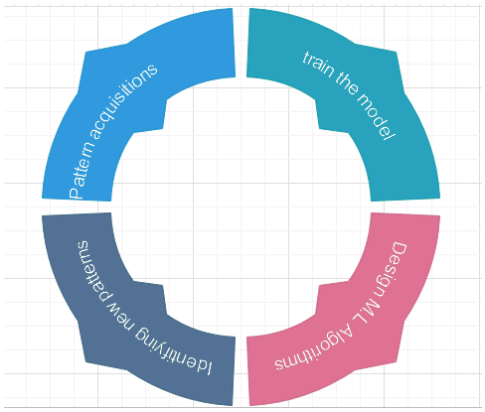

Fig. 1 M.L model steps

a) Pattern acquisitions

Data will be collected which defines how the model could be. The quality of the model completely depends upon how the data is collected.

***b)* Train the model**

Dispute data and make it ready for training. This model trains the data based on the M.L algorithm chosen.

***c)* Design M.L Algorithms**

M.L algorithms are classified into supervised, unsupervised and reinforcement algorithms. Each algorithm is composed of a set of algorithms that are specially designated to solve various problem domains.

***d)* Identifying new patterns**

Identifying the new patterns by the trained model with the help of M.L algorithms and the parameters.

***e)* Validate and Apply**

Evaluate the model by various data patterns and classify according to their parameters. The evaluated model can be applied to various problem domains.

***f)*Proper decision making**

Machine learning is best suited to consider inputs and obtain the appropriate decision termed as a cognitive expert decision system. By means of patterns acquired and experience gained by the experts helps to achieve the decision making.

C. R Features

R is a special-purpose programming language best suited for data analysis, manipulation, visualisation, process and training data. It is a high language that provides a statistical environment and graphical representations to the data scientists to implement various operations on data [1, 2]. The R language is best suited for programmers, statisticians and analysts to take their users to the expert level. R language is rich in data sets, functions, models, plots, graphs and many other utilities. Due to increasing research constants demands, R to be most proficient language to be opted by the statisticians and data scientists [19] [23]. A detailed description of how R language is implemented to improve the classification performance is demonstrated in the subsequent sections.

II. LITERATURE REVIEW

As its fundamental view, data science is the study that aims to make effective use of scientific methods on data et al. (Mujthaba, 2020). These scientific methods include the use of information technology, exhibiting and obtaining business decisions. The future growth of any organization purely depends on the existing data analysis, modelling, processing and presentation [Mujthaba, 2020). Data science is the field that acquaintances with the data scientists to perform their operations on data effectively, which can predict the nearer business decisions.

Machine learning denotes a cluster of techniques that are widely used by data scientists to allow computers to work in an automated way or with minimal interaction of human beings. The machine learning algorithms are best suited for the transformation of learning parameters to appropriate decisions based on data acquainted et al. (Snoek, Jasper, 2012). These authors also focus that tuning these parameters into valuable decisions requires proper guidance of experts, suitable learning mechanisms, right thumb rules & standard procedures and a few brute force techniques. Following these valuable decisions can lead to generating automated procedures which minimize the human level interactions. These parameters demand to development of machine learning algorithms ineffective way to deal with various problem domains. The study of recognizing learning parameters that suits a particular domain is another important concern (GOUHENG ZHAO, 2020). The important role of the researchers is to have a significant study of data acquisition, data modelling, processing and presentations, which is possible by proper machine learning algorithms used by data scientists. There are various machine learning algorithms that are mainly classified into supervised, unsupervised and reinforcement algorithms. Each type is clustered with numerous algorithms like regression, decision trees, support vector machines, logistic representations, appropriate and random forest algorithms. Each algorithm has its own virtues and limitations. The researchers need to have a complete and detailed study of these algorithms to find which is closer to problem domains. Suitable data selection of various domains helps the researchers to find out the risks associated with learning algorithms et al. (Fraser King,2020). The expatriations needed in developing machine algorithms like keen knowledge on statistics, data cleaning, data visualizations & modelling techniques, statistical methods, analytical approaches, problem-solving through programming, databases connectivity and so on et al. (Mujthaba,2020). During the study of complex systems behaviour, it is important to keep in mind the geographical surroundings under which the system is operated (Zouhair Elamrani,2020). The environment under which the system operates will have much impact on the learning parameters of machine algorithms. These researchers have demonstrated behavioural analysis of drivers with a conceptual approach under various geographical surroundings. As human behaviour is not constant while

operating the same system for various times, it becomes very hard to determine the learning mechanisms of machine algorithms. To address all the problems of a system, appropriate algorithms need to be devised under experts view. Supervised algorithms address those problems whose output is dependent upon known independent variables, whereas unsupervised algorithms on unknown variables et al. (Sunil ray, 2017). These algorithms help us to automate the machine to consider data as input, visualize, process and make better decisions in terms of outputs. As these researchers address the major machine learning algorithms among which we select a few to address our problem like regression techniques, decision trees and logistic representations. Amongst various objectives of machine learning, one is to achieve the performance of the system by experiences et al. (Satya D,2020). The experience can be gained by feeding the data as the input to the generic algorithm and developing logic to act upon it. Various fields use machine learning algorithms effectively, among which the health care system is one et al. (Sathya D,2020). Almost in all organizations, data evolving a big role to tackle critical situations and coming up with the best possible decision makings to implement such things. We need the best toolset al. (Luengo, 2020). A computer programming language like R can be annoying or joyful, which provides defies to their learner's et al. (Aphalo, 2020). These researchers also stated that R programming languages stand steadily in implementing the challenges of machine learning in an effective way. R is currently and extensively used by many organizations data scientists to come up with the best decisions. There are many advantages that make the R programming language to be stronger, like user-defined recursive procedures, data frames, data sets, incorporated set of tools, open-source, cross-platform support and many others. The other additional features like tough graphical capabilities, vast visualizations, a broad selection of packages, comprehensive environments, statistical approaches, distributed computing and databases connections. As big research is contributed by R programming in implementing machine learning algorithms due to its object-oriented, statistical computing, best suited for linear, non-linear, integer, binary, and quadratic problems et al. (Lotfi,2020). The supervised machine learning algorithms are known to harvest the general patterns with the assumption of externally given instances as input to predict the future outcomes et al. (Amanpreet Singh,2020). These algorithms are best for classifying the data from the previous instances and producing the output. These algorithms are frequently used by a data scientist to carry out the operations on data efficiently. As the researchers advised

Many fruitful strategies for problem-solving through rule-based & logic-based techniques. Machine learning algorithms are the best fit to correct models for grouping, categorizations and prediction of outputs with the help of logistic representation, support vector machines, random and forest Gaussian Naive Bayes et al. (Manjula C. Belavagi,2016). The most popular supervised algorithms are regression techniques, which includes logistic representations ridge & lasso regression, which are used to find accurate estimations et al. (Benjamin Y. Gravesteijn,2020). Over the past years, results have been based on binary decisions multi-value decisions, which results from relationships among decision trees et al. (Alsolami, Fawaz,2020). Another form of regression is logistic regressions which are based on a statistical model which form logical functions.

III. MATERIALS AND METHODS

In the past years learning environment has played a vital role in the development of the educational field. The impression of effective learning in the classroom will have a direct impact on providing quality education. The emerging trend and enthusiasm in the students will make the learning environment to be interactive between students and teachers. Overall much research is going on towards bringing more interactive tools which help inappropriate learning. Machine learning and Artificial intelligence can be very much valuable to strengthen teaching abilities. Tools developed with Machine Learning and Artificial Intelligence can be useful to enhance teaching capabilities. It can collaborate both teacher and student with respect to learning, modified interactive environment, improved assessment procedures and others. Educational activities need to be more accurate as it is directly related to students and teachers in providing quality education. In our problem domain of the classroom, we demonstrate how machine learning algorithms can be more useful, especially in higher education. With the increasing demand for quality education, the learning environment has become very rigorous. We conduct an experimental study on the students of prince sattam bin Abdul-Aziz University, department of computer science students, to demonstrate how machine learning algorithms can be very much beneficial. To implement this empirical study, we use the R programming language as an interactive environment. A detailed study of the R programming language will be shown in the subsequent sections. In this paper, we highlight how supervised learning like regression, decision trees and logistic representations can be applied to classroom environments by using R implementations.

IV. R IMPLEMENTATIONS

Due to statistical platform, R programming got lost of features to ease its users to create their own data frames. Here in our case study, we represent the classroom learning to improve its performance in operative mechanisms. To achieve this, we create data frames for the students which represent their data by using R data frames.

A. Data Gathering using Data Frames

Here below is the sample data of students, which can be collected using a data frame in the R programminglanguage. Later on, bellow mentioned queries could be performed on the data.

Table 1. Data frames of the Students and their Marks

```
# Create the data frame1.
stud.data <- data.frame(
  std_id = c (1:5),
  std_name = c("Abdalla","Mohammed", "AbdulRaheman","Mujtaba","Manjur"),
  percentage = c(83.3,75.2,91.0,72.9,84.3) ,
    start_grad_date = as.Date(c("2018-01-01", "2019-09-23", "2017-11-15", "2015-05-11", "2016-03-27")),
  stringsAsFactors = FALSE
)
# Print the data frame.
        print(stud.data)
```

```
stud.marks <- data.frame(
  emp_id = c (1:5),
std_name = c("Abdalla","Mohammed","Abdul","Mujtaba","Manjur"),
  computermarks = c(62,82,72,85,95),
  mathematicsmarks = c(63.3,51.9,61.0,72.0,84.25),
  start_date = as.Date(c("2018-01-01", "2019-09-23", "2017-11-15", "2015-05-11","2016-03-27")),
  stringsAsFactors = FALSE
)
# Print the data frame.
print(emp.data)
```

Table 2. Generated tables of Students and their Marks

```
print(stud.data)
std_id     std_name percentage start_grad_date
     1      Abdalla       83.3      2018-01-01
     2     Mohammed       75.2      2019-09-23
     3 AbdulRaheman       91.0      2017-11-15
     4      Mujtaba       72.9      2015-05-11
     5       Manjur       84.3      2016-03-27
print(stud.marks)
emp_id std_name computermarks mathematicsmarks start_date
     1  Abdalla            62            63.30 2018-01-01
     2 Mohammed            82            51.90 2019-09-23
     3    Abdul            72            61.00 2017-11-15
     4  Mujtaba            85            72.00 2015-05-11
     5   Manjur            95            84.25 2016-03-27
```

B. Queries to be Performed over Data Frames like

- Maintaining databases related to students
- Insertion, deletion and modification to data frames
- Calculations like mean, median, mode, quartiles,
- Find the students by their respective percentages
- Extract, expand, add, delete, modify row and column to data frames
- Selecting sunset of data frames
- Get the structure and summary of data frames.

C. Operations on data frames with R Implementations

The following are the operations along with their format using R implementations

Table 3. Data Frames and their Operations

Operation	Format
1. Structure of data frame	str(stud.data)
2. Summary of the data frame	summary(stud.data)
3. print the names of the columns in the data frames	names(stud.data)
4. Select particular rows of a data frame	stud.data[1:2,]
5. Data editor to create the data in excel form (Shown In table 2)	myTable = edit(myTable)
6. Extract selected columns and rows data from the data frames	data.frame(stud.data$std_name,stud.data$start_grad_date)
7. Expand the rows and columns to existing data frames	stud.data$dept <- c("Computers","Mathematics","Computers","Arabic","Engineering")

D. Regression Algorithm

Because R is a statistical programming platform, it's got some pretty elegant ways to extract statistical summaries from data. To extract a few basic stats from a data frame, use the summary () function:

a) Regression Algorithms on Classroom

Linear regression is the technique that maps the relationship between the X and Y variables, i.e. input(x) is mapped with output(y). Similar all the components of x variable like input (x1, x2, x3...Xn) will be mapped with y variable like output (y1, y2, y3…yn). Each element or component of x variable is assigned with specific weights such that (B).y=B0+B1* x. We have types of regressions like simple regression, which is further divided into the independence of observation, normality and linearity

Table 4. Students Data frame summary

```
> summary(stud.data)
     std_id     std_name            percentage     start_grad_date
 Min.   :1   Length:5           Min.   :72.90   Min.   :2015-05-11
 1st Qu.:2   Class :character   1st Qu.:75.20   1st Qu.:2016-03-27
 Median :3   Mode  :character   Median :83.30   Median :2017-11-15
 Mean   :3                      Mean   :81.34   Mean   :2017-06-09
 3rd Qu.:4                      3rd Qu.:84.30   3rd Qu.:2018-01-01
 Max.   :5                      Max.   :91.00   Max.   :2019-09-23
>
```

1) Independence of Observation

Here, as there is only one independent and dependent variable no need to check the relationships among this.

2) *Normality*

To validate whether the dependent variable tracks a normal distribution, use the hist() function hist(stud.data$percentage).

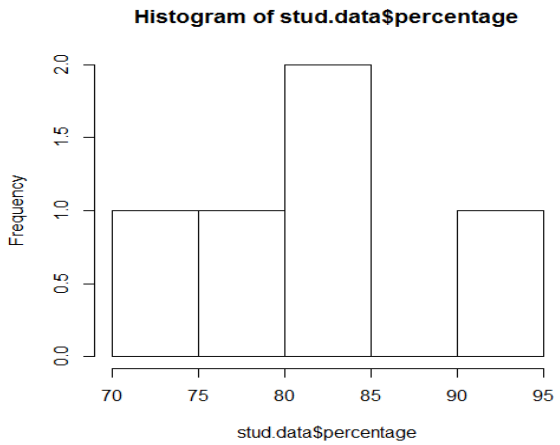

Fig. 2 Students Data Histogram

This figure 2 function is used to plot the graph using hist function between the parameters of class students like their percentage or any numeric field.

3) *Linearity*

The relationship among the dependent and independent variables must exist as linear. We can validate this visual by a scatter plot to see if the distribution of data points could be presented with a straight line

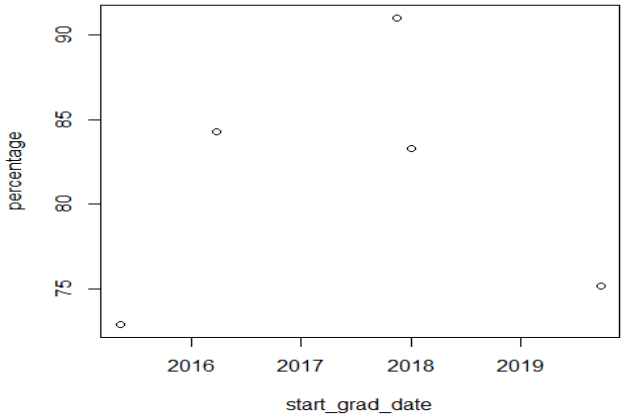

Fig. 3 Linearity between student graduation date and percentages

b) Analysis of Linear regression

To achieve a simple linear regression analysis and pattern the outputs, you need to run a summary of the stud. Data. It is an equation that describes a line that best fits the relationship between the input variables (x) and the output variables (y) by finding specific weightings for the input variables called coefficients (B). It is laidback to implement and effective to train the data. Overfitting of data on X & Y variables can be eliminated with the help of dimensionality reduction, regularization, and cross-validation. The disadvantage of linear regression is that it is suitable for only linear data, not random or dynamic data**.**
summary(percentage.lm)

Table 5. Linear Regression of Students Data

```
Call:
lm(formula = percentage ~ start_grad_date, data = stud.data)

Residuals:
      1       2       3       4       5
 1.703  -7.185   9.462  -7.489   3.509

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     5.968e+01  1.177e+02   0.507    0.647
start_grad_date 1.250e-03  6.788e-03   0.184    0.866

Residual standard error: 8.415 on 3 degrees of freedom
Multiple R-squared:  0.01118,   Adjusted R-squared:  -0.3184
F-statistic: 0.03393 on 1 and 3 DF,  p-value: 0.8656
```

E. Multiple Regression

It is suitable for data sets with multiple numerical values and looks inside the correlation among these variables. This technique is suitable for correlating the dependent variables with independent variables to improve the performance of models. The big advantage of this is to select one variable over another.

a) Independence of observations

Use the stud. data frame to test the relationship between your independent variables, which are closely related

Table 6. Summary of Studentds marks.

```
> summary(stud.marks)
     emp_id     std_name         computermarks  mathematicsmarks
 Min.   :1   Length:5           Min.   :62.0   Min.   :51.90
 1st Qu.:2   Class :character   1st Qu.:72.0   1st Qu.:61.00
 Median :3   Mode  :character   Median :82.0   Median :63.30
 Mean   :3                      Mean   :79.2   Mean   :66.49
 3rd Qu.:4                      3rd Qu.:85.0   3rd Qu.:72.00
 Max.   :5                      Max.   :95.0   Max.   :84.25
   start_date
 Min.   :2015-05-11
 1st Qu.:2016-03-27
 Median :2017-11-15
 Mean   :2017-06-09
 3rd Qu.:2018-01-01
 Max.   :2019-09-23
```

1) Normality

hist(stud.marks$mathematicsmarks)

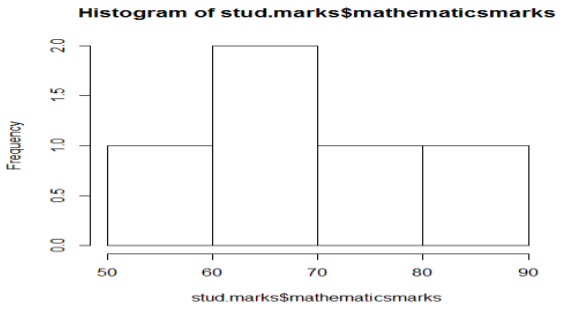

Fig.4 Histogram of the stud. marks data frame using multiple regressions plot(computer marks ~ mathematics marks, data = stud. marks)

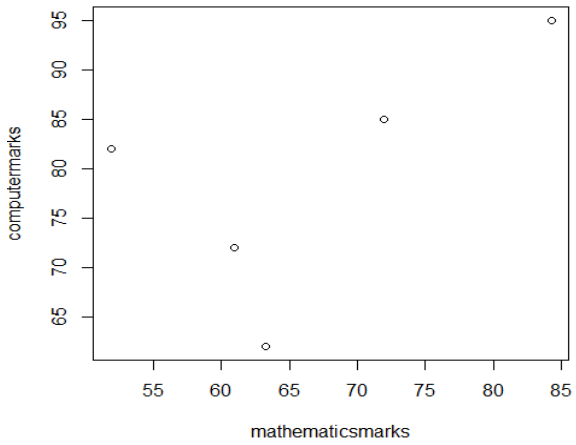

Fig. 5 Multiple Regression plots graphs between computer and mathematics marks

F. Analysis of Multiple regression

This technique is best suited to examine the relationships between dependent and independent variables. The impact of this feature is to analyze, correlate and predict the relationships among different sectors and groups of data.

G. Decision tree

A decision tree is sketched from top to bottom with its root at the top and child in the bottom. Based on the decision made at the root node, the tree splits into left and right internal nodes by means of edges. Each track from top to bottom can lead to a series of suggestions that can lead to a final decision. A complete set of all possible decisions in a tree can form a full decision tree when any internal node is said to be a decision node if it splits to further decision, otherwise known as a terminal node. The question arises of how it can be helpful in maintaining the data sets of students in predicting with right decisions.

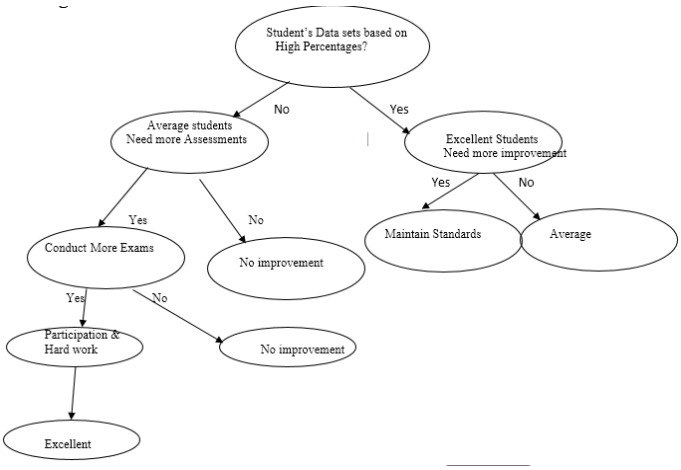

Fig. 6 Decision tree for improving classroom performance

The above figure 6 of a decision tree is used to maintain the data sets of the students divided into two categories. Initially, the root node will make the decision and leads to further decisions. Each internal decision will make further decisions and continue to expansion of the tree from top to bottom. Wherever any node is unable to process further, more decisions are said to a terminal node. If we collectively form and place all the decisions put to gather will generate the final decision.

H. Logistic representations

Logistic regression is another method introduced by machine learning from the area of statistics. It's a prevailing statistical method of modelling a binomial outcome with one or more descriptive variables. It procedures the relationship between the categorical dependent variable and one or more independent variables by evaluating probabilities using a logistic function, which is the cumulative logistic distribution.

V. CONCLUSION

In this research paper, we explicitly demonstrated how machine learning could be useful for systems or software agents to train by themselves rather than programming. Machine learning is rich in algorithms, namely supervised, unsupervised and reinforcement algorithms. Each algorithm is broadly classified into various types. Each algorithm has its own benefits and drawbacks depending upon the application we choose. In our study, we implement machine learning supervised algorithms, namely linear regression, multiple regression, decision trees, and logistic representations, on the application of students' performance in higher studies. Our M.L algorithms, especially supervised algorithms, works on the data frames of students & marks to perform the operations like data gathering, data maintenance, validation, process, interpretation and decision making. All these operations of machine learning are implemented by using an R programming language. The vast applicability and robust nature of R make effective implementations of machine learning algorithms. In the coming research, we recommend implementing other algorithms like unsupervised algorithms on unlabeled data to train the system.

REFERENCES

[1] Lotfi, Farhad Hosseinzadeh, et al. Introductions and Definitions of R. Data Envelopment Analysis with R. Springer, Cham, (2020) 19-52.

[2] Mujthaba, G. M., Al Ameen, A., Kolhar, M., & Rahmath, M. Data Science Techniques, Tools and Predictions.

[3] Forth, Katharine, and Erez Lieberman Aiden. Identifying fall risk using machine learning algorithms. U.S. Patent No. 10,542,914 (2020)

[4] Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. Advances in neural information processing systems. (2012).

[5] Zhao, Gouheng, et al. A Machine Learning-Based Framework for Identifying Influential Nodes in Complex Networks. IEEE Access 8 (2020) 65462-65471.

[6] King, Fraser, et al. Application of Machine Learning Techniques for Regional Bias Correction of SWE Estimates in Ontario, Canada. Hydrology and Earth System Sciences Discussions (2020) 1-26.

[7] Elassad, Zouhair Elamrani Abou, et al. The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. Engineering Applications of Artificial Intelligence 87 (2020).
[8] Biswas, Aniruddha, et al. A Critical Approach to R Programming in the Analysis of lncRNA in Bioinformatics Study. Available at SSRN 3526024 (2020).
[9] Zuccolotto, Paola, and Marica Manisera. Basketball Data Science: With Applications in R. CRC Press, (2020).
[10] Kim, Yeonuk, et al. CH4 flux Gap-Filling Approaches for Eddy Covariance data: a Comparison of three Machine Learning Algorithms and Marginal Distribution Sampling Method with and without Principal Component Analysis. Geophysical Research Abstracts. 21(2019).
[11] Sathya, D., V. Sudha, and D. Jagadeesan. Application of Machine Learning Techniques in Healthcare. Handbook of Research on Applications and Implementations of Machine Learning Techniques. IGI Global, (2020) 289-304.
[12] Luengo, Julián, et al. Big Data: Technologies and Tools. Big Data Preprocessing. Springer, Cham, (2020) 15-43.
[13] Aphalo, Pedro J. Learn R. (2020).
[14] Chinnamgari, Sunil Kumar. R Machine Learning Projects: Implement supervised, unsupervised, and reinforcement learning techniques using R Packt Publishing Ltd, 3(5) (2019).
[15] Singh, Amanpreet, Narina Thakur, and Aakanksha Sharma. A review of supervised machine learning algorithms. 3rd International Conference on Computing for Sustainable Global Development (INDIACom). (2016)
[16] Belavagi, Manjula C., and Balachandra Muniyal. Performance Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection. Procedia Computer Science 89 (2016) 117-123.
[17] Gravesteijn, Benjamin Y., et al. Machine Learning Algorithms Performed no Better than Regression Models for Prognostication in Traumatic Brain injury Journal of Clinical Epidemiology (2020).
[18] Alsolami, Fawaz, et al. Preliminary Results for Decision and Inhibitory Trees, Tests, Rules, and Rule Systems. Decision and Inhibitory Trees and Rules for Decision Tables with Many-valued Decisions. Springer, Cham, (2020) 45-73.
[19] Asthana, Pallavi, and Bramah Hazela. Applications of Machine Learning in Improving Learning Environment. Multimedia Big Data Computing for IoT Applications. Springer, Singapore, (2020) 417-433.
[20] Flach, Peter. Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press, (2012).
[21] Sharon Machlis. Beginner's guide to R: Easy ways to do basic data analysis, Executive Editor, Data & Analytics, Computerworld | (2017).
[22] Peyakunta Bhargavi, Machine Learning Algorithms in Big data Analytics, International Journal of Computer sciences and Engineering ·(2018).
[23] Dr S. Subatra Devi Big Data - Benefits and its Growth International Journal of Computer Trends and Technology 68 (5) (2020) 14-17.
[24] Sangeeta. K, G.V.S.S.Naveen Babu, Madhuri. G Classification and Prediction of Slow Learners Using Machine Learning Algorithms. International Journal of Computer Trends and Technology 68(2) (2020) 54-58.