International Journal of Recent Trends in Engineering & Research
ISSN (ONLINE) : 2455 – 1457
IMPACT FACTOR : 4.101

Machine Pedagogy - Classification and Statistical Analysis of Educational Data

Dr Neha Kumari

Department of Computer Science, St Stephens College, Delhi University, Delhi , India

Abstract— Machine Learning is the science to train the computer systems learning from the past experiences in the form of Data. Machine Learning and Artificial Intelligence can have significant impact in our Education System . This paper discusses how classification Machine learning algorithms paves the way for building a Machine Learning Model for further enhancing the interaction between teachers and students. Statistical test - Pearson Bivariate Correlation , Linear Regression helped in further predictions in making Learning Model . The model prediction have been compared on the basis of performance measures Mean absolute error , Mean Squared error , Gradient Descent.

Keywords— Classification , Regression , Information and Communication Technology , Train set , Test Set , Probablity.

Nomenclature

ICT - Information and Communication Technology
ML - Machine Learning
MAP - Maximum Probability

I. INTRODUCTION

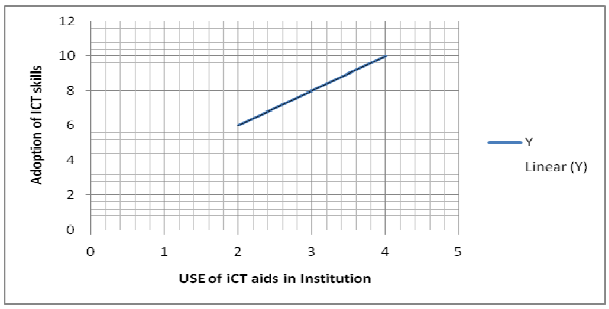
Machine Learning basically signifies a computer program learning from the experiences recorded as data . Alan Turing in his paper stated that interaction of machines with humans is essential for enhancing intellectual temperament of machines. Machine Learning Algorithms can be categorized into Supervised Learning , Semi Supervised Learning, Reinforcement Learning , Decision Trees.
Information and Communication Technology(ICT) has emerged as a major tool for research and innovations permeating every aspect of life. ICT has unfolded plethora of opportunities in IT industry, administrative jobs, teleworking offering copious amount of opportunities for young people. ICT has the potential to nurture reflective thinking of students and cultivate future innovations. Incorporation of technology into the classroom pedagogy needs holistic approach to enhance learning.

II. CLASSIFICATION ALGORITHMS

The six most common classification algorithms were used to classify them among the two classes - understanding the impact of ICT on learning and pedagogies. Schools data with the facilities of projector , power point presentations and students response towards ICT was used to train the classifier.

The study emphasized supervised learning as the desired output was already known. Major classification algorithms used were discussed below :
Linear Regression :- Linear Regression is a Linear model that predicts the value of Y with linear Combination of the values of X. The form of the model would be

$$y = A_o + A_1(x)$$



Naive Bayes's Theorem
Baye's theorem classification algorithm is used when we try to predict the probability of a hypothesis on the basis of our prior knowledge.
P(h|d) = (P(d|h) * P(h)) / P(d)
Where
P(h|d) is the probability of hypothesis h given the data d. This is called the posterior probability.

- **P(d|h)** is the probability of data d given that the hypothesis h was true.
- **P(h)** is the probability of hypothesis h being true (regardless of the data). This is called the prior probability of h.
- **P(d)** is the probability of the data .

After calculating probability from various hypothesis . We choose the one with the maximum probability.
MAP(h) = max((P(d|h) * P(h)) / P(d)).
With the help of Naive Bayes's theorem , the probability of the

III. SAMPLING AND DATA ANALYSIS

The research was exploratory in nature . In total there were 53 statistical variables. Cluster and quota sampling was used to sample the data. Statistical test used to draw inferences from the data - Bivariate Correlation and Anova. Data was collected from total 441 respondents . The data sets was first divided into training data sets and test data sets.

TRAIN SET
The Train set for developing a model included 200 observations without any missing data and proper values for predicting the model.

TEST SET
The test set included remaining values that were not taken in the training set.

Bivariate Pearson Correlation Test was applied between Use of ICT aids in institution and adoption of ICT skills as described in the table below:

		Use of ICT tools in Institution
Use of ICT tools in institution	Pearson Correlation	1
	Sig. (2-tailed)	
	N	441
Email	Pearson Correlation	.736**
	Sig. (2-tailed)	.000
	N	441

Learning from educational resources available online	Pearson Correlation	.769**
	Sig. (2-tailed)	.000
	N	441
Using social networking websites for professional networking	Pearson Correlation	.869**
	Sig. (2-tailed)	.000
	N	441
	N	441
Understanding of the subject	Pearson Correlation	.724**
	Sig. (2-tailed)	.001
	N	441
Retention of facts and figures	Pearson Correlation	.853**
	Sig. (2-tailed)	.001
	N	441
Classroom presentations make classes more interactive.	Pearson Correlation	.739**
	Sig. (2-tailed)	.002
	N	441

It is noticeable from the table that the correlation was statistically significant at .01 level i.e. (99% level of confidence). Since the Pearson Correlation Coefficient r is positive, it indicates positive correlation. That is if one variable increases, the value of another variable also tends to increase and if one variable decreases, the value of other variable also tends to decrease.

If Pearson coefficient r value is close to 1, it indicates strong relationship between the two variables and if Pearson coefficient value approaches 0, it indicates weaker relationship between the variables.

The table indicates positive correlation. Use of ICT tools in institution is positively correlated with adoption of technical skills. In all the cases value of Pearson Correlation coefficient r approaches 1 indicating strong relationship between the two variables.

IV. CONCLUSION

1. Bivariate Correlation indicated positive relation between the prediction and responses.
2. Linear Regression and Naive Bayes 's theorem classified the data into classes . It can be implied that use of technology as an ICT aid helps in better retention of facts and figures , clear understanding .

REFERENCES

I. Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996
II. [2] M. Chen, Z. Xu, K. Weinberger, and F. Sha. *Marginalized denoising autoencoders for domain adaptation*. In Proceedings of the 29th International Conference on MachineLearning, pages 767{774. ACM, 2012.
III. Alaa el-Halees, "Mining *students data to analyze e-Learning behavior: A Case Study*", 2009.
IV. S. T. Hijazi, and R. S. M. M. Naqvi, "*Factors affecting student"s performance: A Case of Private Colleges*", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.
V. Jordi Bieger, Kristinn R. Thórisson, and Bas R Steunebrink. 2017. *The pedagogical pentagon: conceptual framework for artificial pedagogy.* International Conference on Artificial Intelligence (Lecture Notes in Computer Science, vol (10414), Tom Everitt, Ben Goertzel, and Alexey Potapov.

