

# Data Science and Machine learning: A Comparative Analysis for Integration of Digital Literacy Competences into Teacher Education Program using Machine learning Models in Rwanda Secondary school

Niyonshuti Eugene <sup>1</sup> Master's, Dr. Musoni Wilson<sup>2</sup> PhD, Uwamahoro Bonaventure <sup>2</sup>  
Master's, Kanyandekwe Emmanuel <sup>2</sup> Master's.

<sup>1,2</sup> Information communication and Technology (ICT), Kigali, Rwanda.

Subject area: Machine Learning

**Abstract:-** This article investigated and explored a Comparative analysis for integration of Digital literacy competences into teacher Education Program using Machine learning Models in secondary school, The global objective of this research was to make a comparative analysis for integration of Digital literacy competences into teacher Education Program using Machine learning Models in Rwanda education system and the impact that technology may have as a key performance on teaching and learning system. The findings confirmed that the teachers used Digital literacy competences in their teaching program in 2020 was lower Because most of teachers were no trained on Digital literacy integration in education system which is indicated by 2%. The simulation has indicated that all teachers were trained in 2020. in 2021 the finds show as that all teachers are using Digital literacy competences in their teaching program which is indicated by 100%. The study also recommends using this study of analysis for integration of digital literacy competences into teacher education program.

## I. INTRODUCTION

Quality education is a key concern as a means of preparing learners for the knowledge economy. Many countries now regard understanding Digital literacy and getting to know the primary abilities and concepts of Digital literacy integration as part of the Key performance in education. One of UNESCO's overriding aims is to ensure that all countries, both developed and developing have access to the best educational facilities necessary to prepare young people to play full roles in modern society and to contribute to a knowledge-based nation.

For ICT integration in education to thrive, adequate preparation is required. It necessitates careful planning, excellent teacher preparation, ongoing teacher professional support, and visionary leadership that sees the need of preparing students to live and work in the technology world of the twenty-first century. Rwanda Vision 2020 aims at moving Rwanda from “an agriculture-based economy to a knowledge-based society “and middle-income country 2020, Education is a key sector to this social and economic

transformation, tapping into the limitless potential of an empowered population, at the same time the Vision 2020 places ICTs at the heart of the transformation across all sectors (Rwanda, 2016). See digital literacy's contribution to education. proposed research is aiming at investigating the integration of digital literacy and its contribution in teaching and learning as a key performance in Rwanda secondary schools by examining how familiar teachers and students are with ICT facilities, the relevant technology and its applications, the research will continue by examining the nature of the Impact of digital literacy Integration on Secondary School Education and Learning Processes, this will imply Teacher involvement who are trained in digital literacy in ngoma district.

## II. METHODOLOGY

### A. Data Analysis

Analysis of data is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, in different business, science, and social science domains. Data mining is a particular data analysis technique that focuses on modeling and knowledge discovery for predictive rather than purely descriptive purposes.

Sales Forecasting is considered Supervised Machine Learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.  $Y = f(X)$ , The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stops when the algorithm achieves an acceptable level of performance

**B. Cleaning the Data**

Cleaning data should be the first step in any Data Science (DS) or Machine Learning (ML) workflow. Without clean data it'll be much harder time seeing the actual important parts in the exploration. Once the training of ML models begins, they'll be unnecessarily more challenging to train. The main point is that to get the most out of a dataset, it should be clean. In the context of data science and machine learning, data cleaning means filtering and modifying the data such that it is easier to explore, understand, and model. *Filtering out* the parts you don't want or need so that you don't need to look at or process them. *Modifying* the parts, you *do need* but aren't in the format you need them to be in so that you can properly use them. The dataset used needed the following changes to be considered clean:

- Dropping of Rows with Null Values
- Removal of all negative numbers and replacing them with absolute values

**C. Null Values**

Most data science algorithms do not tolerate nulls (missing values). So, one must do something to eliminate them, before or while analyzing a data set. There are many techniques for handling nulls. Which techniques are appropriate for a given variable can depend strongly on the algorithms you intend to use, as well as statistical patterns in the raw data, in particular, the missing values, and the randomness of the locations of the missing values. Moreover, different techniques may be appropriate for different variables, in a given data set. Sometimes it is useful to apply several techniques to a single variable. Finally, note that corrupt values are generally treated as nulls.

The figure below shows the rows in the dataset used in this article and the number of missing values in each row.

**D. Data Visualization**

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

**A. Interpretation of the results**

- a) Digital literacy integration skills and knowledge in the teaching-learning process

Year	Teacher_use_DL_Prep_Lesson
2020	3 of 103
2021	103of 103

Table 1: Teacher uses Digital literacy in preparing lesson

**III. RESULT**

- **Machine learning models:** we divided our dataset into two variables X as the features we defined earlier and y as the Product\_Branch\_Sale the target value we want to predict. This is a regression problem so we used Regression methods. Train test split will be ratio respectively.
- **Machine Learning Models used:**
  - Linear Regression
  - Random Forest Regressor
  - Lasso Regressor
  - Gradient Boosting Regressor
  - Decision Tree Regressor
  - Ridge Regressor
- **Deep Learning Model used:** Artificial Neural Network
- **The Process of Modeling the Data:**
  - Importing the model
  - Fitting the model
  - Predicting Product Sales
  - Regression metrics
  - Score Metrics for Regression: Mean Absolute Error (MAE) - Mean of the absolute value of errors (absolute distance from true value):
  - Mean Squared Error (MSE) - Mean of the squared value of errors (squared distance from true value):
  - R<sup>2</sup> (coefficient of determination) - Regression score function.
- **Linear Regression:** In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.
- Mean Absolute Error: 1.2588781740944013e+17
- Mean Squared Error: 1.5455189915440857e+37
- R<sup>2</sup> score: -2.963116350531758e+25

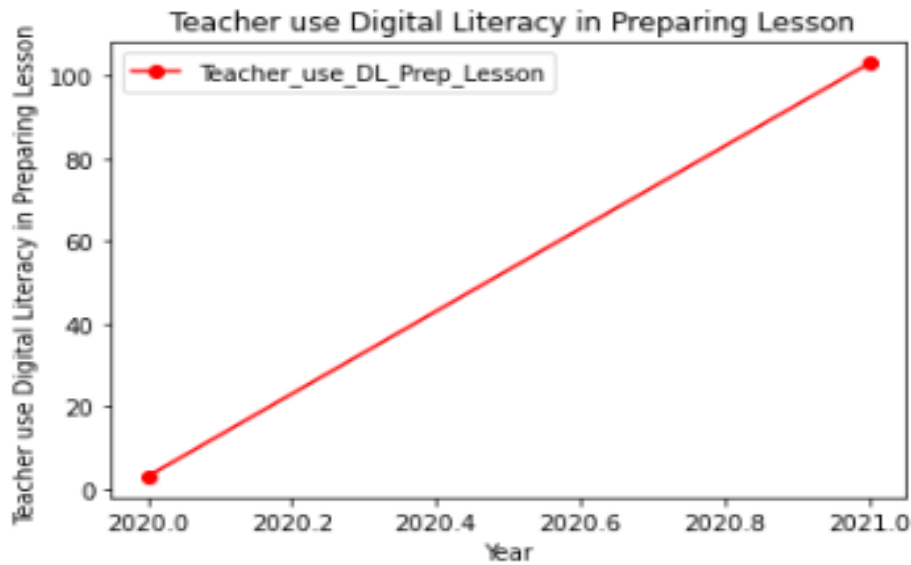


Fig. 1: Linear Regression graph

Results in figure 1 revealed that the use of digital literacy in preparing lessons for teacher in 2020 was lower or almost inexistent before teachers were given trainings. This is indicated by only 3 out of 103 teachers which are 3% who used this system for teaching in 2020. In 2021, many teachers were trained to use digital literacy in their teaching-learning process, which is explained by 103 out of 103 teachers which are 100% who were using the digital literacy in their teaching-learning process.

Random forest is a Supervised Learning algorithm which uses ensemble learning method for classification and regression. It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Mean Absolute Error: 81005.29  
 Mean Squared Error: 18076378196.53  
 R<sup>2</sup> score: 0.9653

b) Random Forest Regressor

Year	Teacher_Strategies_teach_learn_process
2020	1 of 103
2021	102 of 103

Table 2: Teacher Strategies Teacher Learn Process



Fig. 2: Strategies of teaching-learning process are relying on Digital literacy Source: Secondary data, 2020-2021

Results in figure 6 have shown that the use of strategies of teaching-learning process are relying on Digital literacy was lower or almost inexistent in 2020, before teachers were given trainings. During this period, teachers were obliged to work from home in form of e-Learning or online learning where digital literacy was the only option to avoid physical contact with students. This is indicated by only 1 out of 103 teachers who used this system for teaching.

In 2021, many teachers were trained to use digital literacy in their teaching-learning process, which is explained by 102 out of 103 teachers who were using the digital literacy in their teaching learning process. This

clearly indicates that strategies of teaching-learning process are relying on Digital literacy.

c) Lasso Regressor

In statistics and machine learning, lasso (least absolute shrinkage and selection operator; also Lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces

Mean Absolute Error: 204450.06  
 Mean Squared Error: 86313641129.96  
 R^2 Score: 0.8345

B. Teachers' D.L integration as key performance in teaching-learning process

Year	Teacher_use_DL_in_admin_work
2020	2 of 103
2021	103 of 103

Table 3: Teacher use digital literacy in administrative work and evaluation process

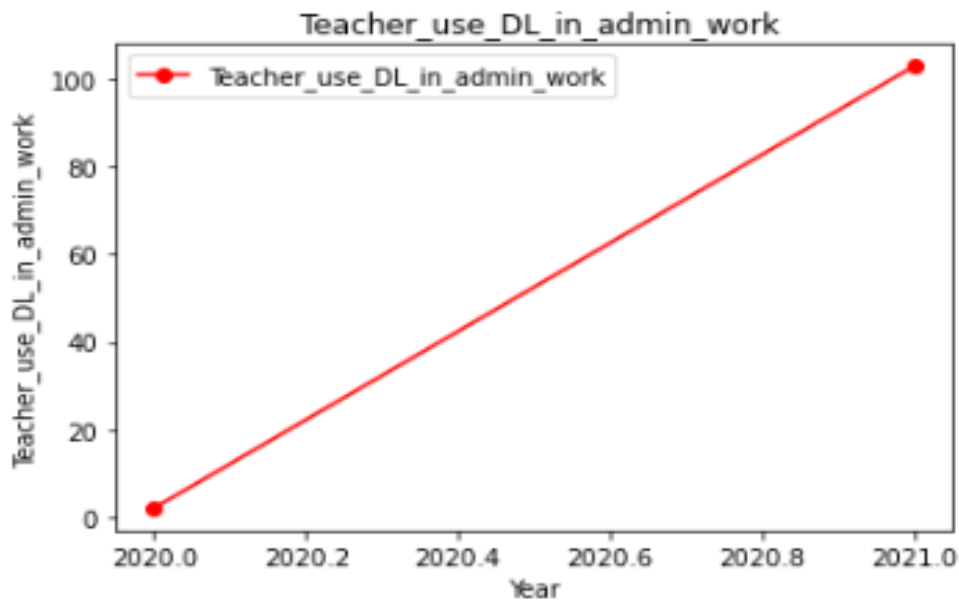


Fig. 3: Teacher use D.L in administrative work & evaluation process

Source: Secondary data, 2020-2021

The figure 7 has shown that Teacher use digital learning in administrative work & evaluation process, which is 100% relying on Digital literacy, but the number of teachers who were using it was too small or almost inexistent in 2020, before teachers were given trainings. This is indicated by only 2 out of 103 teachers who used this system for teaching. In 2021, many teachers were trained to use digital literacy in their teaching-learning process, which is explained by 103 out of 103 teachers who were using the digital literacy in their teaching-learning process.

This clearly indicates that teachers use digital learning in administrative work & evaluation process. During this period, teachers were obliged to work from home in form of e-Learning or online learning and services were provided online, where digital literacy was the only option to avoid physical contact with students.

a) Decision Tree Regressor

Mean Absolute Error: 97020.48  
 Mean Squared Error: 21931465357.44  
 R^2 Score: 0.958

Year	Teacher_strategy_of_teaching_depend_on_Digital_literacy
2020	1 of 103
2021	103 of 103

Table 4: Teacher Strategy of Teaching Depend on Digital literacy

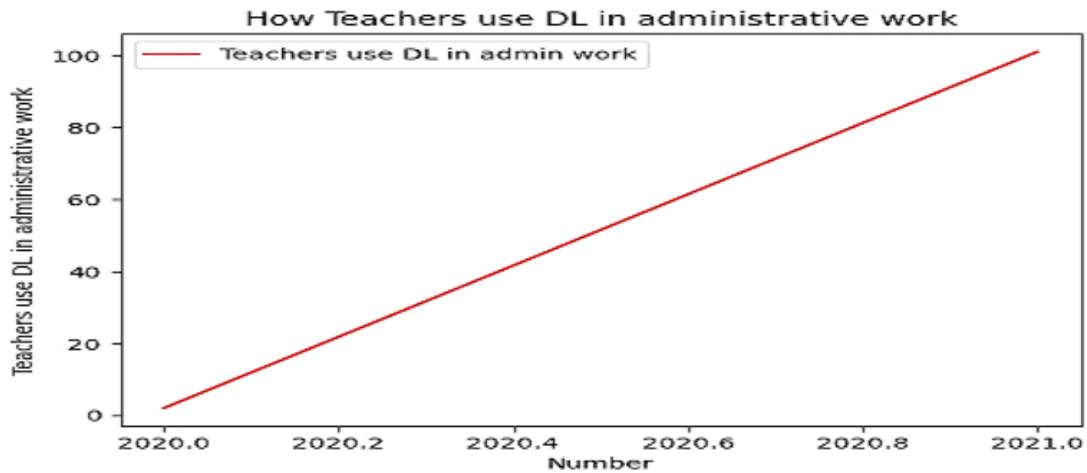


Fig. 4: Teacher Digital strategy of teaching depend on Digital literacy Source: Secondary data, 2020-2021

The figure 8 has shown that teacher digital strategy of teaching depends on Digital literacy, but the number of teachers who were using it was too small or almost inexistent in 2020, before teachers were given trainings. This is indicated by only 1 out of 103 teachers who used this system for teaching. In 2021, many teachers were trained to use digital literacy in their teaching-learning process, which is explained by 102 out of 103 teachers who were using the digital literacy in their teaching-learning process. This

clearly indicates that teachers use digital learning in administrative work & evaluation process teacher digital strategy of teaching depend on Digital literacy.

- b) Ridge Regressor  
 Mean Absolute Error: 189059.25  
 Mean Squared Error: 72037890447.05  
 R<sup>2</sup> Score: 0.8619

C. Digital literacy in preparing teaching materials

Teacher use Digital literacy in preparing teaching materials

Teacher_use_DL_prep_teach_materials	Number
2020	1 out of 103
2021	103 out of 103

Table 5: Teacher\_use\_DL\_prepare\_teacher\_materials

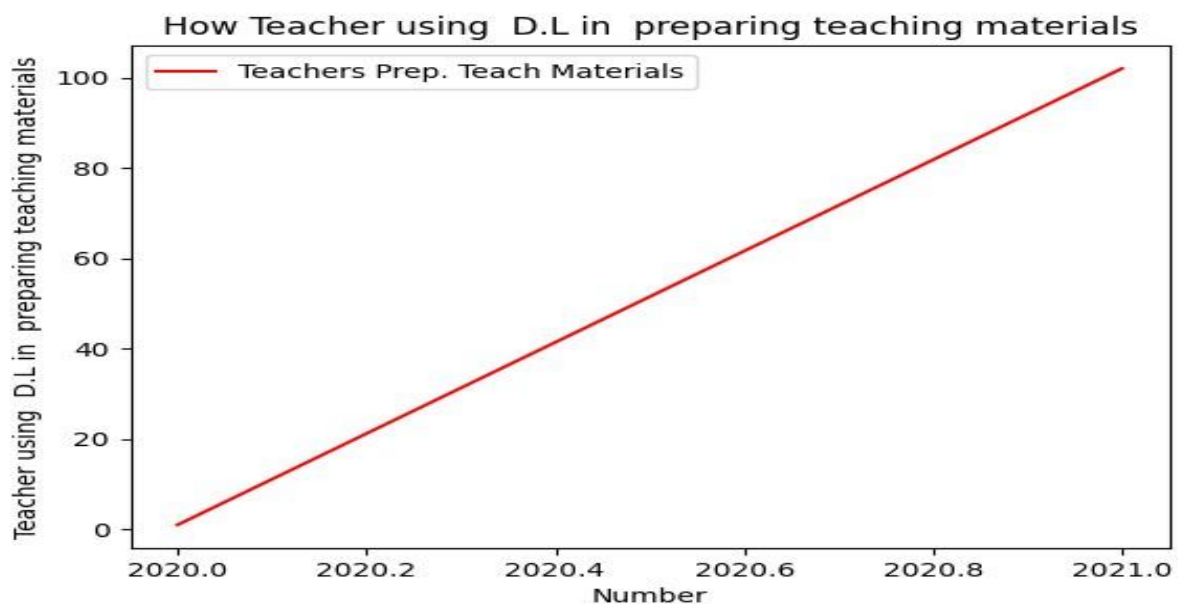


Fig. 5: Teacher use Digital literacy in preparing teaching materials

Source: Secondary data, 2020-2021

The information from figure 9 show that teacher use digital literacy in preparing teaching materials, which is 100% relying on Digital literacy, but the number of teachers who were using it was too small or almost inexistent in 2020, before teachers were given trainings. This is indicated by only 1 out of 103 teachers who used this system for teaching. In 2021, many teachers were trained to use digital literacy in their teaching-learning process, which is explained by 103 out of 103 teachers who were using the digital literacy in their teaching-learning process after being given training on the use of digital literacy in their teaching-

learning process. This clearly indicates that Teacher use Digital literacy in preparing teaching materials. During this period, teachers were obliged to work from home in form of e-Learning or online learning and materials were provided online, where digital literacy was the only option to avoid physical contact with students.

- a) Gradient Boosting Regressor  
 Mean Absolute Error: 21281.12  
 Mean Squared Error: 882118510.38  
 R<sup>2</sup> score: 0.9983

Teacher_use_AudioVisual_teach_learn	Number
2020	1 out of 103
2021	103 out of 103

Table 6: Teacher use Audio –Visual teaching learning

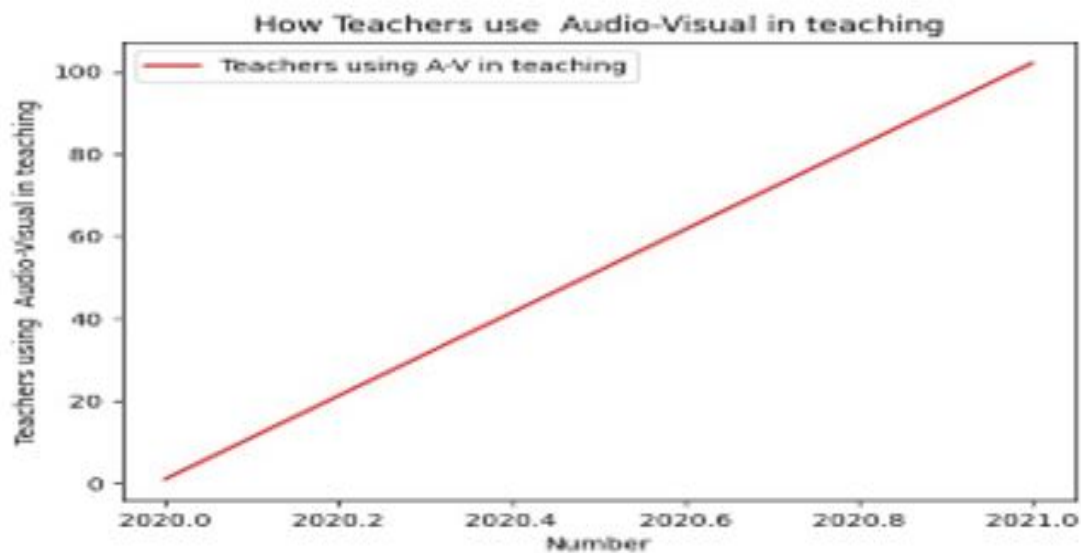


Fig. 6: Teacher use Audio –Visual teaching learning

Source: Secondary data, 2020-2021

The information provided by Figure 10 show that teachers use audio–visual teaching learning. Before 2021, the number of teachers who were using it was too small or almost inexistent, because teachers were not yet given trainings.

This is indicated by only 1 out of 103 teachers who used this system for teaching. In 2021, many teachers were

trained to use digital literacy in their teaching learning process, which is explained by 102 out of 103 teachers who were using the digital literacy in their teaching-learning process after being given training on the use of digital literacy in their teaching-learning process. This clearly indicates that teachers use audio–visual teaching learning.

Teacher_use_as_resources_teach_doc	Number
2020	2 out of 103
2021	103 out of 103

Table 7: Teacher use Digital literacy as resources of teaching document

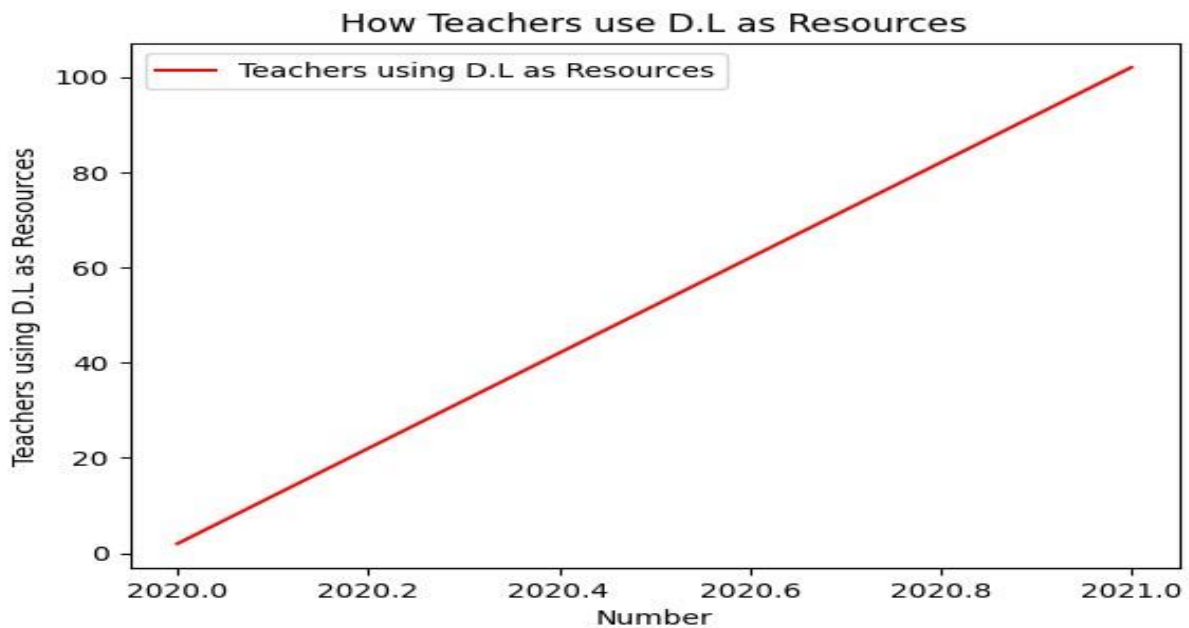


Fig. 7: Teacher use digital literacy as resources of teaching document

Source: Secondary data, 2020-2021

The information provided by Figure 11 show that teachers use digital literacy as resources of teaching document. Before 2021, the number of teachers who were using it was too small or almost inexistent, because teachers were not yet given trainings. This is indicated by only 1 out of 103 teachers who used this system for teaching. Many teachers were trained to use digital literacy in their teaching-

learning process, which is explained by 103 out of 103 teachers who were using the digital literacy in their teaching-learning process after being given training on the use of digital literacy in their teaching-learning process. This clearly indicates that teachers use digital literacy as resources of teaching document.

b) Multilayer Perceptions

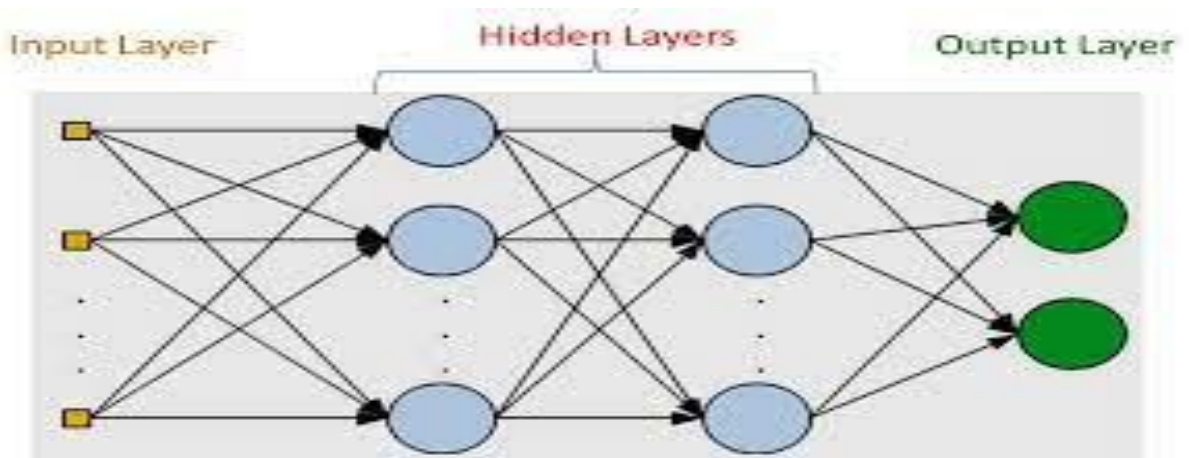


Fig. 8: Architecture graph of MLP

The MLP model stands for Multilayer Perceptron is a type of feed-forward artificial neural network (ANN) where the information flows from the input layer towards the output layer through the hidden layer. Rectified Linear Unit is used as activation function for Multilayer Perceptron algorithm. MLP makes use of a supervised learning algorithm called backpropagation for training the network. In backpropagation, the error is propagated backward throughout the network. The error is calculated by taking

the difference between the network output and the actual output. The network parameters called weights are modified to minimize this error based on this method. This process is repeated several times until a stopping condition is reached.

Mean Absolute Error: 51421.09  
 Mean Squared Error: 11623311059.29  
 R<sup>2</sup> score: 0.9777

This Article makes use of the GBR model stands for Gradient Boosting Regressor is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. If a small change in the prediction for a case causes no change in error, then next target outcome of the case is zero

The GBR model is well suited for this article because of the following reasons:

- GBR is suitable for this article because it classified prediction problems where inputs are assigned a class or label
- GBR method is used to forecast the sales revenue of upcoming period. According to results there are high similarities between forecasted and actual data.
- GBR is suitable for regression prediction problems where a real-valued quantity is predicted given set of inputs

#### IV. CONCLUSION

The main objective of this article was investigated and explored a Comparative analysis for integration of Digital literacy competences into teacher Education Program using Machine learning Models in secondary school, The global objective of this research was to make a comparative analysis for integration of Digital literacy competences into teacher Education Program using Machine learning Models in Rwanda education system and the impact that technology may have as a key performance on teaching and learning system. In 2021, many teachers were trained to use digital literacy in their teaching-learning process, in contrast to 2020 where almost no teachers were trained. The findings of this research confirmed that the integration of digital literacy has the greater impact on teaching learning process, here the number of teachers who were using it was too small or almost inexistent in 2020, before teachers were given trainings. This is indicated by only 3 out of 103 teachers who used this system for teaching in 2020.

And the find show that in 2021 almost 100% teachers relying on digital literacy in teacher education program 2021. This is indicated by only 102 out of 103 teachers who used this system for teaching in 2021. In conclusion the integration of digital literacy has the greater impact on teaching learning process.

#### REFERENCES

- [1.] Busch, P. A. (2017). The role of contextual factors in the influence of ICT on street-level discretion. *Proceedings of the Annual Hawaii International Conference on System Sciences, 2017-Janua*, 2963–2972. <https://doi.org/10.24251/hicss.2017.358>
- [2.] Daniels, J. (2008). Information and Communication Technology in Education: Opportunities. *Journal of Baltic Science Education*, 5(5), 5–7.
- [3.] Ghavifekr, S., & Rosdy, W. A. W. (2015). Teaching and learning with technology: Effectiveness of ICT integration in schools. *International Journal of Research in Education and Science*, 1(2), 175–191. <https://doi.org/10.21890/ijres.23596>

- [4.] Minishi-majanja, M. K. (2007). Date: 07/06/2007 Integration of ICTs in Library and Information Science Education in sub-Saharan Africa Mabel K. Minishi-Majanja. *Library and Information Science*, 1–19.
- [5.] Ministry of Youth and ICT (MYICT). (2015). *ICT Sector Profile - 2015*. 38.
- [7.] Mubaraka, C. M., Senyonga, T., Sumil, N. R., Principal, D., & Lecturer, S. (2013). *Effectiveness of a Proposed System Design on Academic Management of Kampala International University: Design, Development and Implementation*. 4(5), 12–16.
- [8.] Wisler, A. K. (2009). 'Of, by, and for are not merely prepositions': teaching and learning Conflict Resolution for a democratic, global citizenry. *Intercultural Education*, 20(2), 127–133. <https://doi.org/10.1080/14675980902922143>
- [9.] <https://docs.aws.amazon.com/machine-learning/latest/dg/training-ml-models.html>
- [10.] <https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
- [11.] [https://en.wikipedia.org/wiki/Linear\\_regression](https://en.wikipedia.org/wiki/Linear_regression)