Framework for the Development of an Enhanced Machine Learning Algorithm for Non-Cognitive Variables Influencing Students' Performance using Feature Extraction

Oguns Yetunde Josephine[1], Ayeni Joshua Ayobami[2] and Ganiyu Aminat Abidemi[3]
[1]Lecturer, Department of Computer Science, The Polytechnic, Ibadan, Oyo State, Nigeria
[2]Associate Professor, Department of Computer Sciences, Ajayi Crowther University, Oyo, Nigeria
[3]Lecturer. Department of Computer Science, The Polytechnic, Ibadan, Oyo State, Nigeria

[1]Corresponding Author: oguns.yetunde@polyiadan.edu.ng

Received: 28-06-2023 Revised: 13-07-2023 Accepted: 28-07-2023

ABSTRACT
Machine learning is a powerful tool for creating computational models in scientific analysis in areas where there is need to extract hidden data such as educational data. In order to make planning easier and identify at-risk students who may be in danger of failing or dropping out of school due to their academic performance, Educational Data Mining (EDM) uses computational tools. In this paper, a framework using machine learning approach was proposed to develop an enhanced algorithm for non-cognitive variables influencing students' performance using feature extraction. In the framework, the Decision Tree (DT) and Linear Support Vector Machine (SVM) are proposed as base classifiers, and Random Forest (RF) and Gradient Boosting (GB) as ensemble classifiers. The DT classifier allows the classification process to be modelled as a series of hierarchical decisions on the features, forming a tree-like structure. Using this technique, planning and predicting students who might be at-risk of dropping out would have been made easier.

Keywords: hierarchical, educational, classifiers, non-cognitive, dropping

I. INTRODUCTION

Machine learning (ML) is a set of techniques that, in general, allow us to "train" computers how to execute tasks by showing them how to do them. ML is a field that enables computers to make successful predictions using past experiences. It is the process of programming computers to maximize a performance criterion based on previous experience or example data. Models of ML establishes parameters, that learning through the process of running a computer program to optimize the model's parameters using training data or prior experience. The model may be predictive to make future predictions, or descriptive to learn from data, or both. The application of ML methods to large databases is called data mining. In data mining, a vast amount of data is analyzed in order to create a basic model with useful properties, such as having high predictive accuracy. Data mining is a multidisciplinary field that permits one to obtain relevant information from large amounts of data at the confluence among other areas: artificial intelligence, statistics, databases, and information science. The purpose of this technique is to discover previously unknown patterns. Once these patterns have been discovered, they may be used to make specific decisions about how to grow an organization, business, or enterprise [1].

In recent years, there's been a surge in interest in using data mining to study scientific questions in educational research, a field known as Educational Data Mining (Baker, 2010). Educational data mining is a branch of science concerned with the development of methods for making discoveries from the particular types of data generated in educational settings and employing such methodologies to gain a deeper understanding of students and the environments in which they learn. In explicitly exploiting the multiple levels of meaningful hierarchy in educational data, educational data mining methods frequently differ from methods in the broader data mining literature. To achieve this goal, methods from the psychometrics literature are frequently combined with methods from the ML and data mining literature. Prediction of students' academic success has long been considered an important study topic in many academic areas because it aids both teaching and learning. Rates of Drop-outs in higher education learning are high. Data mining techniques can be used to harness student data to generate early warnings about students who are in danger of failing and enable appropriate interventions, uncover enrollment habits linked to academic success, and improve school advice and certificate preparation.

II. LITERATURE REVIEW

John McCarthy was one of the founders of the discipline of artificial intelligence. He coined the term “artificial intelligence” (AI) and was very influential in the early development of AI [20]. The mathematician In order to put his theories and queries into practice, Alan Turing was able to test if "machines can think" [21]. After a series of tests, which came to be known as the Turing Test, it is discovered that it is feasible to make robots think and learn like humans [21]. The Turing Test employs a pragmatic approach to determine whether machines can react like humans [3, 4]. In their work, Liu et al. [4] stated that AI has expanded significantly, institutionalized more and more, and broadened study in this period of multidisciplinary science, including computer science, cybernetics, automation, mathematical logic, and linguistics. It is crucial to be ready for AI revelations, exactly as the UAE has done by appointing an AI state minister in Dubai [3].

2.1 Machine Learning

Machine learning is a term that encompasses a number of techniques or methods that enable the investigator to learn from the data. These methods equally allow fast translation to applications of large real-world databases to inform patient-provider decision making. According to [5], the ubiquitous and potent kind of artificial intelligence is currently transforming every business. A kind of artificial intelligence known as machine learning allows computers to learn without being explicitly programmed [5]. In just the last five or 10 years, machine learning has become a critical way, arguably the most important way, most parts of AI are done and that is the reason most researchers use the terms AI ad ML interchangeably or as synonymous [5]. Contrary to computer system commands, which are created to produce an output based on an input (see figure 1), there are some circumstances in which computers make decisions based on the sample data that is now available and in those circumstances, Computers may make mistakes similar to how people make decisions [6]. In other words, machine learning is the act of giving computers the ability to learn from data and experience similarly to how a human brain does [7]. Machine learning's primary goal is to build models that can learn from past data to become better, recognize intricate patterns, and solve new issues [8] (as depicted I Figure 1). Brownlee [10] stated in his work that ML enables computers to generate their own programs. This fundamental principle that underlies the concept of machine learning explains why it is so important in modern computing. Machine learning is distinguished by automating the process of automation, as opposed to conventional programming, which merely reflects automation. Artificial intelligence, which is widely defined as a machine's ability to mimic intelligent human behavior, includes the subfield of machine learning. Artificial intelligence (AI) systems are used to carry out complicated tasks in a manner akin to how people solve issues [22]. Traditional programming involves giving a computer data and a program as input so it can output a certain result, but machine learning uses data and an output to generate a program from start.

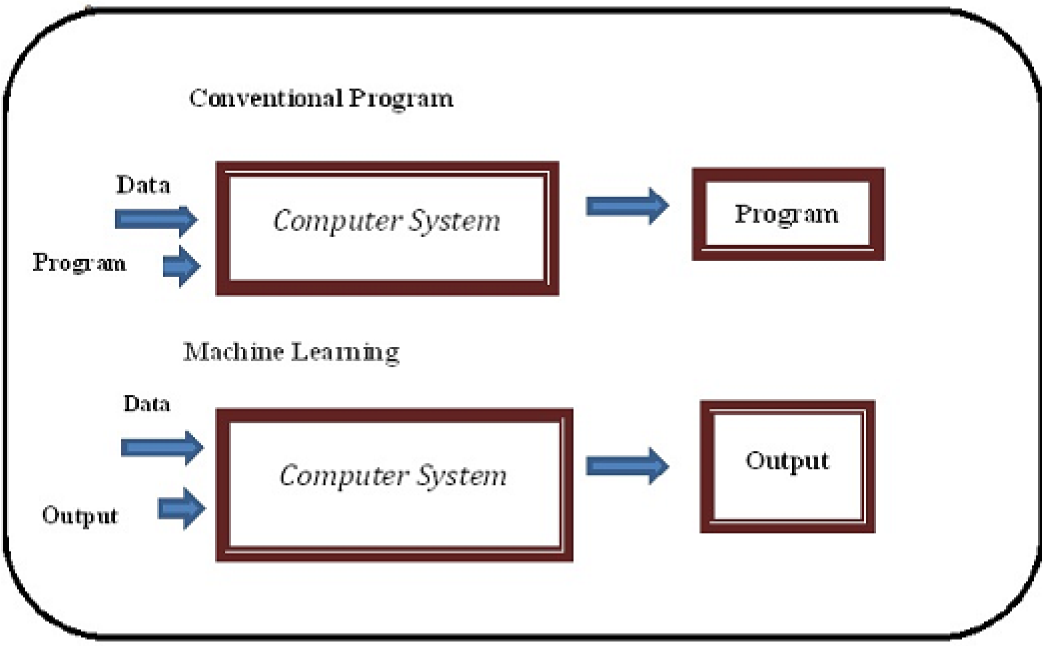

Figure 1: Illustrative diagram of ML and Conventional Programming

2.2 Feature Extraction in Machine Learning

A dimensionality reduction technique called feature extraction divides a large initial set of raw data into smaller groupings that may be processed more easily [11]. These huge data sets share the trait of having many variables that demand a lot of computational power to process. The process of selecting and/or combining variables into features, known as feature extraction, significantly reduces the quantity of data that needs to be processed while precisely and thoroughly defining the initial dataset [11]. Additionally, by minimizing the amount of redundant data, feature extraction can aid analysis (see figure 2). Additionally, the machine's efforts to generate variable combinations (features) and the reduction of the data speed up the learning and generalization phases of the machine learning process [12] (see Figure 2).

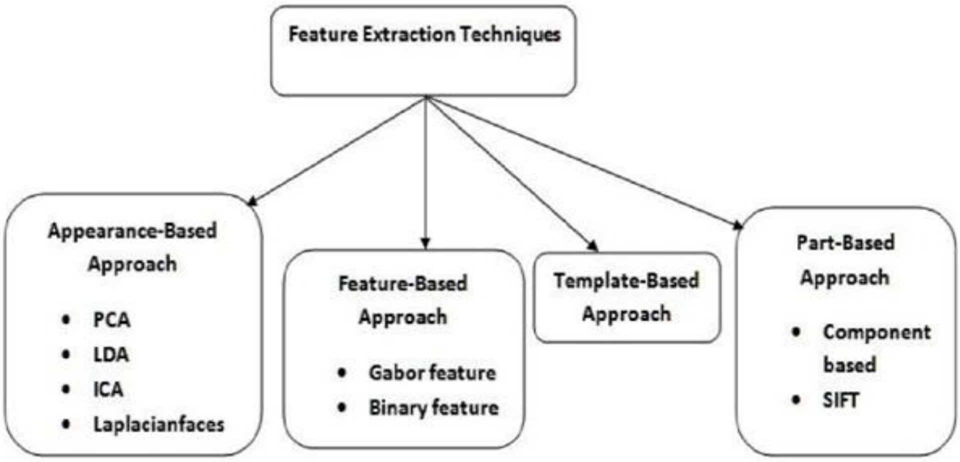

Figure 2: Classification of the various Feature Extraction Techniques
Source: ([15])

By selecting the most important aspects of the data for machine learning, an initial set of data is reduced through the process of feature extraction, which seeks to minimize the number of features in a dataset by constructing new features from the current ones (and then discarding the unnecessary ones) [11],[12]. The process of converting unprocessed raw data into numerical features that can be processed while keeping the details of the original data set could also be used to describe it.

III. MATERIALS AND METHODS

The Data Selection stage is a necessary requirement in this work as most of the past researchers in the field of EDM have often built their datasets used to train the developed students' performance prediction model privately. These datasets are gathered manually via questionnaires or from the information management system utilized by educational institutions. Again, due to the peculiarities of the datasets, which cannot be compared with the features and feature selection methods needed in this work, the students' historical academic dataset from the UCI machine learning repository and Kaggle, which are the two internationally renowned open machine learning data warehouses, could also not be used as the experimental datasets.

In this framework, the first step of the research methodology is the dataset acquisition and was constituted with the dataset of the undergraduate students of the department of Computer Sciences, Ajayi Crowther University, Oyo, ICT and covers a period of two (2) years as an example dataset. The dataset was analyzed and examined if the feature extraction and selection procedure could be appropriately used to find out student data patterns with the use of MsExcel (Microsoft Excel). The main goal at the stage is to clean the dataset and then feed them into the feature extraction modules of the developed students' performance prediction model and generate some sort of pattern, suitable for the ML model; Supervised (Classification vs Regression) or Unsupervised (Clustering vs. Dimension Reduction vs. Association). The various steps of the process is show in Figure 3.

3.1 Identification of Significant Features in the Dataset

The levels of study for a program as contained in dataset the dataset are classified according to the student's level of study and an instance of a student in a course refers to his/her performance at the end of a given semester in that course. Courses taken before the past semesters are referred to as previous courses for lower level; Prerequisite courses are courses that must be taken by student before being allowed to take a dependent course.

Instance (I):= P, S, c. (i)

Cs:pv := Previous courses taken by Student S past semesters
Cs:x := Set of Courses for a given semester
Cs:cr := Credits allocated to a given course courses
PreqCs, Csq := Prerequisite courses

e-ISSN: 2583-553X
Peer Reviewed & Refereed Journal
https://doi.org/10.31033/abjar.2.4.6

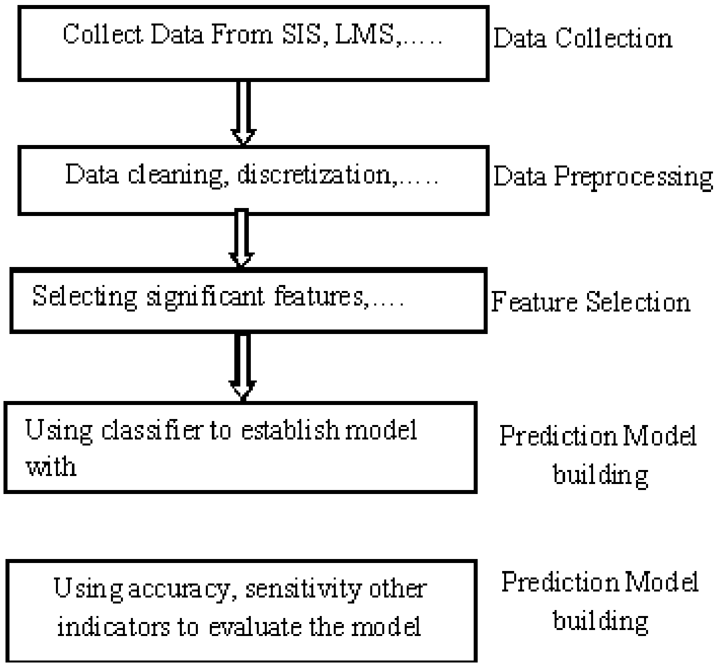

Figure 3: Representation of the Prediction of students' performance Processes (Adapted from [13])

a. Grade Point (GP)

The Grade Point derives from the percentage raw score for a given course, the raw score is converted into a letter grade and a grade point as shown in the Tables 1 and 2 [16].

b. Grade Point Average (GPA) [16]

The Grade Point Average (GPA) summarizes achievement for each semester. This is the cumulative average or weighted grade point average for the semester's course work. The Grade Point Average is calculated by multiplying the Grade Point earned in each course by the number of credit units assigned to that course, adding these results together, dividing by the total number of credit units taken for the semester [16].

c. Cumulative Grade Point Average (CGPA)

This represents the current average of the Grade Points the student earned throughout a course of study. The total Grade Points multiplied by the corresponding credit units for each semester are added and then divided by the total number of credit units for all courses the student has registered for to determine the student's overall performance at any time in the B. Sc. Program [16].

Table 1: Scoring and Grading System

Credits I (1-6)	**Present Scores II**	**Letter Grades III**	**Grade Points (GP) IV**	**Grade Point Average (GPA)**	**Cumulative (CGPA**
Vary according to contact hours assigned to each course per semester and according to work load carried by students	70-100	A	5	Derived by multiplying I and IV, sum up for all courses and dividing by total credit units per semester	4.50 - 5.00
	60-69	B	4		3.50 - 4.49
	50-59	C	3		2.40 - 3.49
	45-49	D	2		1.50 - 2.39
	40-44	E	1		< 1.50

Table 2: Grades are awarded with the score in a given course:
(**Source:** www.acu.edu.ng, Student Handbook, Dept. CSC 2023) [16]

Score Range	Grade	Points
70 and above	A	5
60 – 69	B	4
50 – 59	C	3
45 – 49	D	2
40 – 44	E	1

Given the above historical data from the dataset, different features are used in capturing factors that could affect student's performance. Polyzou and Karypis in [14] identified three distinct features;

- The student-specific (independent from course c),
- Course-specific features (independent from student s) and student-
- Course-specific features (they are a function of both s and c).

3.2 Feature Extraction Technique

The Principal Component Analysis (PCA) technique is proposed for feature selection/extraction from the dataset. The PCA algorithm feature extraction/selection is listed following (Algorithm 1): The Sklearn tool of Python (public domain) is also proposed. (Algorithm 1).

Algorithm 1: The PCA algorithm feature extraction/selection

Step 1: Perform one-hot encoding to transform categorical data to numerical data set
Step 2: Perform training / test split of the dataset
Step 3: Standardize the training and test data set
Step 4: Construct covariance matrix of the training data set
Step 5: Select the most important features using explained variance
Step 6: Construct project matrix;
Step 7: Transform the training data set into new feature subspace

Having carried out the cleansing of the data, data preparation, data encoding and data transformation to convert the data into an executable format and enhance model performance, selecting appropriate ML algorithms for better and accurate prediction results is the next stage of the work.

3.3 Select the Appropriate Machine Learning Model to Train the Dataset

The following Python classifiers are suggested for use with the scikit-learn library: Both ensemble classifiers Random Forest (RF) and Gradient Boosting (GB) are used. Machine learning methods called ensemble classifiers integrate multiple base models to create a single, ideal predicting model [17]. For classification, regression, and other tasks, Random Forest is an ensemble learning technique that builds a large number of decision trees during the training phase and outputs the class that represents the mean of the classes (classification) (see Figure 4) or mean prediction (regression) of the individual trees [18]. Gradient Boosting is a machine learning method for classification and regression problems that generates a prediction model in the form of a group of weak prediction models, often decision trees [17].

3.4 Evaluation and Validation Technique

The evaluation of the results of a ML algorithm could be achieved using different techniques and one of the following performance metrics is proposed; accuracy, precision (specificity), recall (sensitivity), and the F1 score. These values are determined from the confusion matrix (Table 3). Accuracy (eqn. 2) is defined as the number of correctly predicted instances over the total number of records, precision is the ratio of correctly predicted positive instances to the total predicted positive instances, sensitivity is calculated as the ratio of the number of correctly predicted instances to the total number of positives, and the F1 score is the weighted average of precision and sensitivity [18].

$$\text{Accuracy} = \frac{\text{TP} + TN}{\text{TP} + FP + TN + FN} \quad (2)$$

e-ISSN: 2583-553X
Peer Reviewed & Refereed Journal
https://doi.org/10.31033/abjar.2.4.6

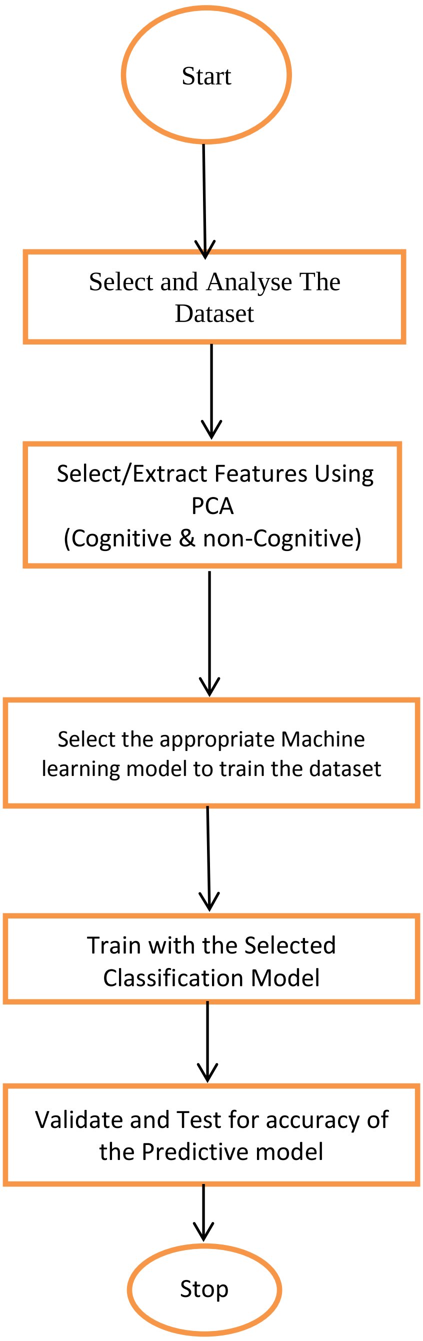

Figure 4: Flowchart for the Proposed Algorithm

Table 3: Confusion matrix [22]

		Predicted Values	
		Positive	*Negative*
Actual values	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Precision = $\frac{TP}{TP+FP}$ (3)

Sensitivity = $\frac{TP}{TP+FN}$ (4)

Precision is defined (eqn. 3) as the ratio of genuine positives to all expected positives. Precision is defined as the ratio of successfully predicted positive classes to all forecasted items. The proportion of true positives to all true positives is known as recall, which is defined as the ratio of correctly predicted positive classes to all true positives. While recall relates to the classifier's ability to locate all of the positive samples, precision refers to the classifier's ability to not label a negative sample as positive. The F1 score (eqn. 5) is a measure of accuracy that is defined as a single performance metric that takes into account both recall and precision. It is also often known as the F-Measure. It is calculated by harmonically averaging the two metrics and calculated as [18].

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision } + \text{Recall}} \tag{5}$$

IV. CONCLUSION

This framework has illustrated the basic methodology for identifying students at-risk of dropping out of school due to poor performance using non-cognitive variables it has also demonstrated the fundamental underlying concept of the use of Machine Learning technics as computational tools to explore educational data for the purpose of identifying hidden data. Intelligent machine systems are being used as a more effective alternative to manual analysis in EDM decision support system due to advances in technology and computational capacity. This framework proposed DT and SVM as base classifiers, while RF and GB as ensemble classifiers in order to improve the accuracy of the prediction model. The development ad adoption of the system will be of great use to the advantage of students and the Nigerian Universities in general. Identifying such students early with an increased attentional attitude (often needed) by the educators and teachers/lecturers will be of advantage to the school, students and parents.

ACKOWLEDGEMENT

The contributions of members of the department of Computer Sciences Ajayi Crowther University, Oyo, Nigeria and the postgraduate students are hereby acknowledged.

REFERENCES

1. Bharati, M., & Ramageri, B. (2010). Data mining techniques and applications. *Indian Journal of Computer Science and Engineering.*
2. Baker & Ryan. (2010). Data mining for education. *International Encyclopedia of Education*. *7*, 112-118.
3. Mariam Khaled, & Alsedrah. (2017). Artificial intelligence. *America University of The Middle East*. doi: 10.13140/RG.2.2.18789.65769. Available at: https://www.researchgate.net/publication/323498156_Artificial_Intelligence.
4. Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., & Lee, I. ((2018). Artificial intelligence in the 21st century. *IEEE Access*, *1*(1), 10.1109.
5. Sara Brown. (2021). Machine learning, explained. Available at: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.
6. Çelik, Ö, & Altunaydın, S.S. (2018). A research on machine learning methods and its applications. *Journal of Educational Technology & Online Learning*, *1*(3), 25-40.
7. Gör, İ. Vektör, Nicemleme İçin, Geometrik Bir, Öğrenme Algoritmasının, & Tasarımı ve Uygulaması. *Yüksek lisans tezi, Adnan Menderes Üniversitesi, Aydın*.
8. Tantuğ, A. C., & ve Türkmenoğlu, C. (2015). Türkçe metinlerde duygu analizi. *Yüksek lisanstezi, İstanbul Teknik Üniversitesi, İstanbul*.
9. Malhotra, P. (2016). *Why machine learning is being given so much Importance?*. Delhi: India.
10. Brownlee, J. (2015). Basic concepts in machine learning. *Retrieved Machine Learning Mastery*.
11. Deep AI. (2022). *Feature extraction*. Available at: https://deepai.org/machine-learning-glossary-and-terms/feature-extraction.
12. Pier, P. I. (2019). Feature extraction techniques. *Published in Towards Data Science*. Available at: https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be.
13. Wen Xiao, Ping Ji, & Juan Hu. (2021). A hybrid feature selection method for predicting students. *Hindawi Scientific Programming*, 16. https://doi.org/10.1155/2021/1670593.
14. Polyzou, A., & Karypis, G. (2018). Feature extraction for classifying students based on their academic performance. *11th International Conference on Educational Data Mining*. EDM , Buffalo, United States.
15. Balan, S., & Sunny, L.E. (2018). Survey on feature extraction techniques in image processing. Available at: https://www.semanticscholar.org/paper/Survey-on-Feature-Extraction-Techniques-in-Image-Balan-Sunny/a5d8205ee67c873f2a2b46373d8e3d8000e00f48.
16. *Student Handbook*. Dept. of Computer Sciences, Available at: www.acu.edu.ng.

17. Pedregosa *et al.* (2011). Scikit-learn: Machine learning in python. *JMLR, 12*, 2825-2830.
18. Jared Wilber. (2022). *Precision & recall: Accuracy is not enough*. Available at: https://mlu-explain.github.io/precision-recall/.
19. Tabea Tietz. (2020). John mccarthy and the raise of artificial intelligence. Available at: scihi.org/john-mccarthy-artificial-intelligence.
20. Brian McGuire. (2006). The history of artificial intelligence, History of computing, CSEP 590A. University of Washington. Available at: https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf.
21. Sara Brown. (2021). *Machine learning, explained*. Available at: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.
22. Karimi, Zohreh. (2021). *Confusion matrix*. Available at: https://www.researchgate.net/publication/355096788_Confusion_Matrix.