iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.59920

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

An Apt Process for Machine Learning and Artificial Intelligence

Dr P Vijaya Vardhan Reddy
SM-IEEE, IEEE, India

Abstract*: The paper presents an apt process for Machine Learning and Artificial Intelligence. To solve the real-world problems with Machine Learning (ML) and Artificial Intelligence (AI) a definite process is required which would emphasis on complete solution rather than approaching as randomly. Machine Learning and Artificial Intelligence are not just mere algorithms which you can put anywhere and start getting fantabulous results. ML and AI are processes which starts with defining the data and completes with the model with defined level of accuracy. In this paper an apt process is proposed to solve the real-world problems with Machine Learning and Artificial Intelligence.
Keywords: Machine Learning, Artificial Intelligence, Process, Data, Model

I. INTRODUCTION

To solve the real-world problems with Machine Learning (ML) and Artificial Intelligence (AI), just having or selecting great algorithms is not sufficient and does not yield solutions to the problem. Machine Learning and Artificial Intelligence are required an apt process to solve the problems in any domain. Machine Learning process starts with defining the data and ends with model. An apt process proposed in this paper for Machine Learning includes defining the problem, collecting the data, preparing the data, splitting the data for training and testing, algorithm selection, training the algorithm, evaluation on test data, parameter tuning, and start using the model. The apt process proposed in this paper for Artificial Intelligence includes defining the problem, data collection and preparation, selecting models and algorithms, training the model, evaluating model performance, fine-tuning, and optimization, deploying the model, and ethical considerations.

II. RELATED WORK

Artificial Intelligence (AI) and especially machine learning (ML) become increasingly more frequently applicable in all operations. This paper presents a systematic review of today's applications of ML techniques in an environment. The utilization of ML methods related to process planning and control, predictive maintenance, quality control, in situ process control and optimization, logistics, robotics, assistance and learning systems for shopfloor employees are being analysed. Moreover, an overview of ML training concepts in learning factories is given [1]. In addition, identified three aspects of the AI domain that make it fundamentally different from prior software application domains: 1) discovering, managing, and versioning the data needed for machine learning applications is much more complex and difficult than other types of software engineering, 2) model customization and model reuse require very different skills than are typically found in software teams, and 3) AI components are more difficult to handle as distinct modules than traditional software components - models may be "entangled" in complex ways and experience non-monotonic error behaviour. The lessons learned by Microsoft teams will be valuable to other organizations [2]. Artificial intelligence (AI) and machine learning (ML) have caused a paradigm shift in healthcare that can be used for decision support and forecasting by exploring medical data [3]. AI engineering that provides an overview of the key engineering challenges surrounding ML [4].

Machine learning and artificial intelligence are becoming the most talked of ML or AI-focused studies in the literature have increased almost exponentially [5]. ML engineering may apply to data-centric AI as the problem of designing data collection, labelling, and quality monitoring processes for datasets to be used in machine learning [6]. The processes AI can enhance or inspire to sharpen competitive edges. the implications of machine learning and AI for both identifying the model and data [7]. AI has led to Machine Learning's development to teach the machine. The "teaching" process happens by providing data from previous operations [8]. AI and ML algorithms are applied to data and can then predict future outcomes via complex processes [9]. AI can be applied to various types of healthcare data (structured and unstructured). Popular AI techniques include machine learning methods [10].

All previous woks have done their research in respective fields and the use of AI and ML has been increased exponentially. In this paper, we present an apt process for ML and AI with step-by-step evaluation to solve the real-world problems irrespective of domain.

©IJRASET: All Rights are Reserved

III. MACHINE LEARNING PROCESS

This section provides an apt process for Machine Learning to solve the anticipated problem, The process starts with defining the problem and ends with the model with defined level of accuracy. Fig. 1 depicts Machine Learning (ML) in nutshell and the section elaborates step by step ML process from Problem Definition to Model.

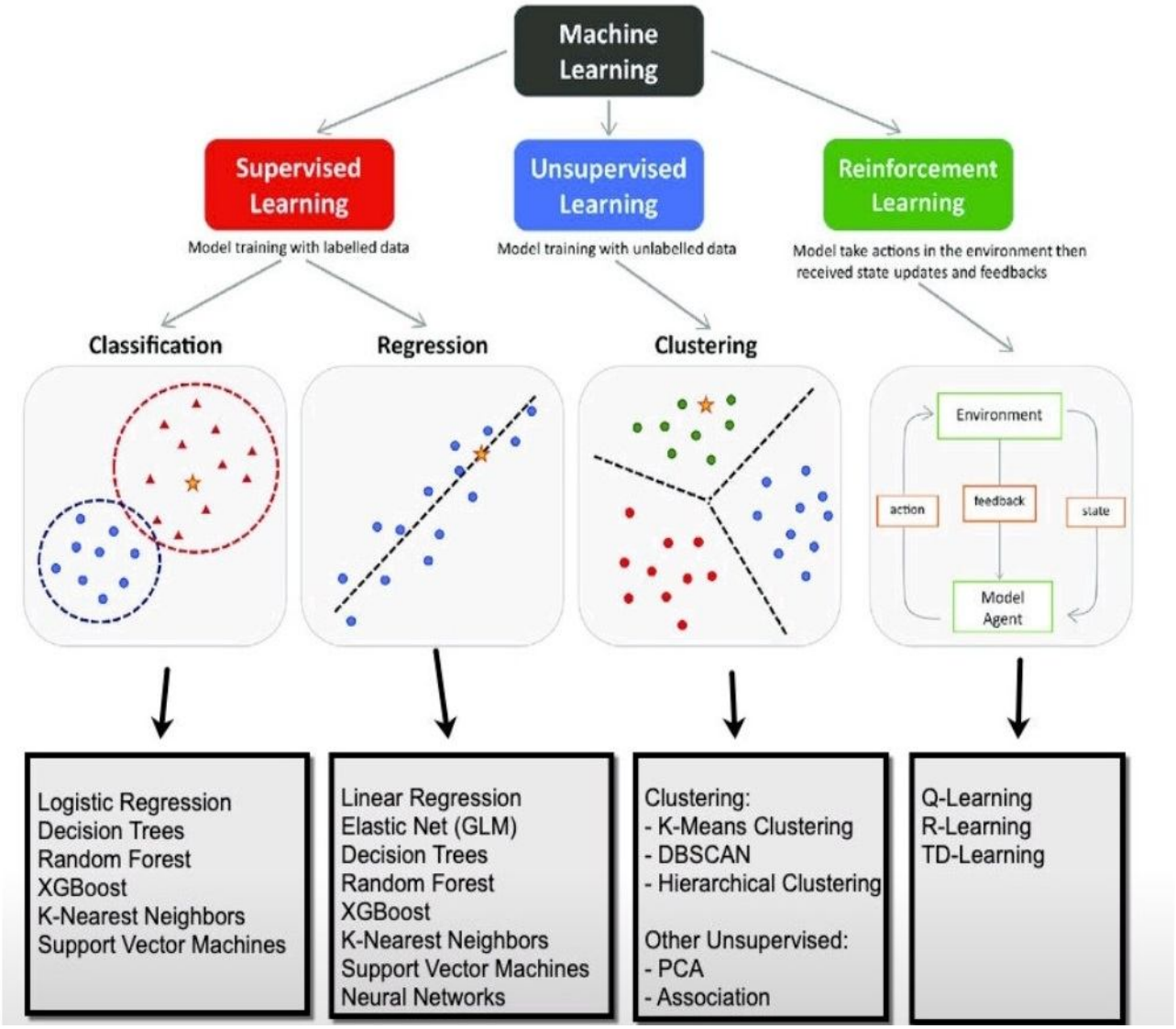


Fig. 1.Machine Learning (ML) in Nutshell

A. The Problem Definition

Definition of business problem consists of two parts; what is the problem and second one is why does this problem need a solution. The definition of problem gives formal context to the real problem. Consider a task to find an image contains human or not. Then to define a problem divide them into 3 parts, as Task (T), Experience (E), and Performance (P). Task is classifying an image contains human or not, Experience is images with label contains human or not and Performance is error rate. Lower error rate leads to high accuracy.

B. Data Collection

Data collections starts after defining the problem. Web scraping and APIs are used to collect the data. Data collection with classification is important. The right data is apt to any machine learning problem. More and better data leads to generate better results even with basic algorithms.

C. Data Preparation

Algorithms does not do any magic or trick. Need to input right form of data to get the results, hence data preparation is the key. Cleaning, formatting, sampling, decomposition, and scaling are the steps to prepare the data aptly.

D. Split the Data in Training and Testing

As per industry standard testing data should be 40 to 20 % and Training data should be 60 to 80 %. The model which gives best results on test data is considered and accepted.

E. Algorithm Selection

The problem definition is the key input to the algorithm selection. In case of classification of emails as spam or not spam, classification algorithms such as Decision Tree, Naïve Bayes, Neural Networks etc. need to be selected. To predict continuous variables regression algorithms such as liner regression or kernel regression need to be opted. No output, but to be grouped them based on properties clustering algorithms need to be selected. So, selection of algorithm is an important step in the proposed apt process to solve the machine learning problems. Fig. 2 lists Machine Learning (ML) algorithms.

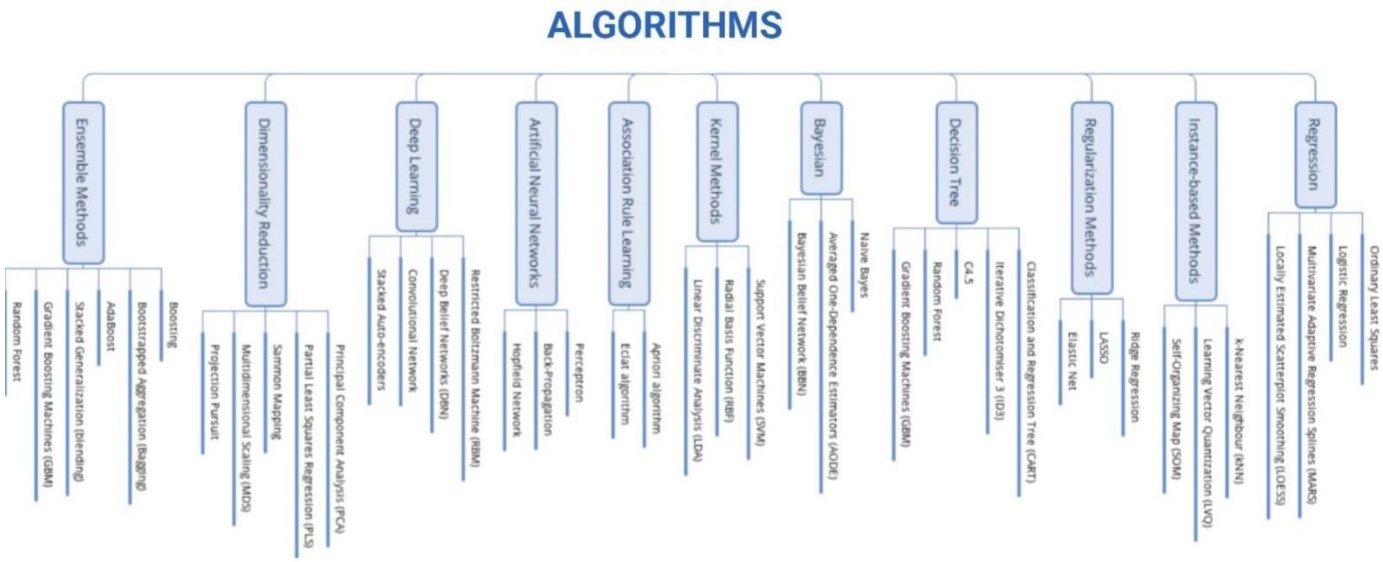


Fig. 2.Machine Learning (ML) Algorithms

F. Algorithm Training:

The algorithms start with random assignment of weights or parameters and improve them in each iteration. In training algorithm, steps run several times on the training dataset to produce results.

G. Test Data Evaluation:

Algorithm decisions are not be biased by test dataset points cos test dataset is not available to algorithm during training. After creating best algorithm on training data, the test data evaluation is performed which is important as Algorithm decisions are not be biased by test dataset points.

H. Tuning of Parameters:

The parameter tuning is each algorithm has different types of settings to configure and change their performance. Rate of learning can be changed to improve the performance. Modifying the parameters is like an art and these parameters are called hyper parameters.

I. Start Using the Model:

The model is ready as it has been trained and evaluated on test dataset. The model can be used and start predicting the values for new data points. The model can be deployed in production server and used its prediction power by communicating using APIs.

This model is not same for all data points. Whenever new data is available start repeating all above steps as mentioned in proposed apt process to improve the performance. Fig. 3 specifies Machine Learning (ML) Models.

Machine Learning (ML) models consists of supervised, unsupervised, ensemble, reinforcement, and neural networks. Supervised models are classification and regression. Unsupervised models are clustering, pattern search and dimensionality reduction. Ensemble models are boosting, bagging, and stacking. Reinforcement models are Q-learning, Deep O-Network (DON), A3C, Generic, and SARSA.

Neural Network models are Multilayer Perception (MLP), Generative Adversarial Networks (GAN), Radial Basis Function Networks, Transformers, Convolutional Neural Networks (CNN), Autoencoders, and Recurrent Neural Networks (RNN). Recurrent Neural Networks (RNN) models further classified into LSTM and GRU. Classification consists of KNN, Logistic Regression, Naïve Bayes, Decision Tree and SVM. Regression consists of linear regression, polynomial regression, Lasso, and Ridge. Clustering contains k-means. DBSCAN, Fuzzy C-Means, Mean-Shift. Pattern search consists of FP-Growth, ECLAT.

Dimensionality Reduction contains PCA, t-SNE, LDA, ODA, LLE, SVD, and LSA. Fig. 3 clearly depicts all mentioned Machine Learning (ML) Models.

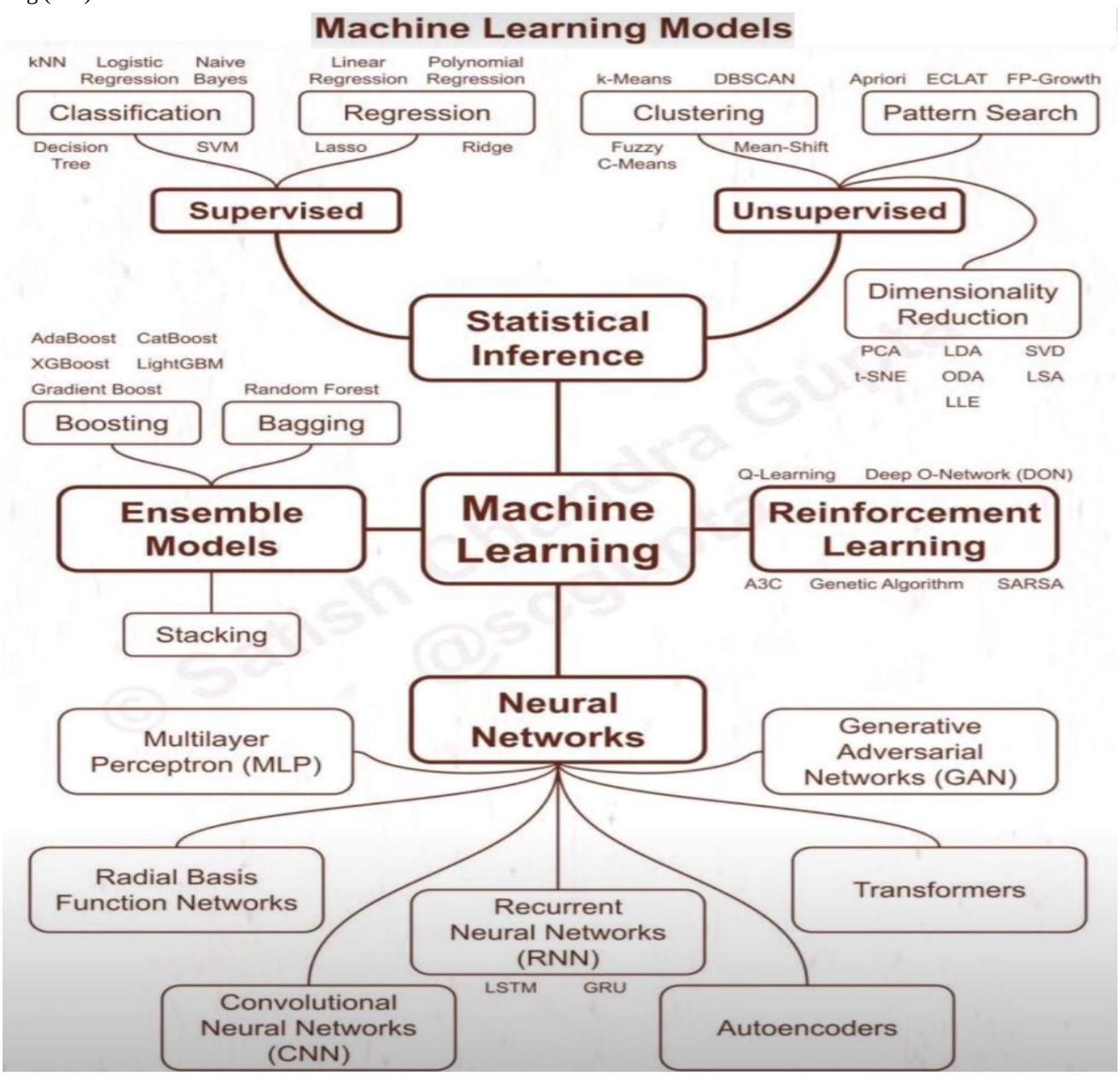


Fig. 3.Machine Learning (ML) Models

IV. ARTIFICIAL INTELLIGENCE PROCESS

This section provides an apt process for Artificial Intelligence (AI) to solve the anticipated problem. Fig. 4 depicts AI in nutshell. AI solution process starts with problem selection, definition, data cleansing and selection, check for bias class representation, oversampling and under-sampling, algorithm selection with classical vs deep learning, model creation with feature selection, weights, hyperparameters, with data distribution in training and testing, accuracy metrics with F-score and ends with monitoring of reproductive results, verifiable, and identify new data elements.

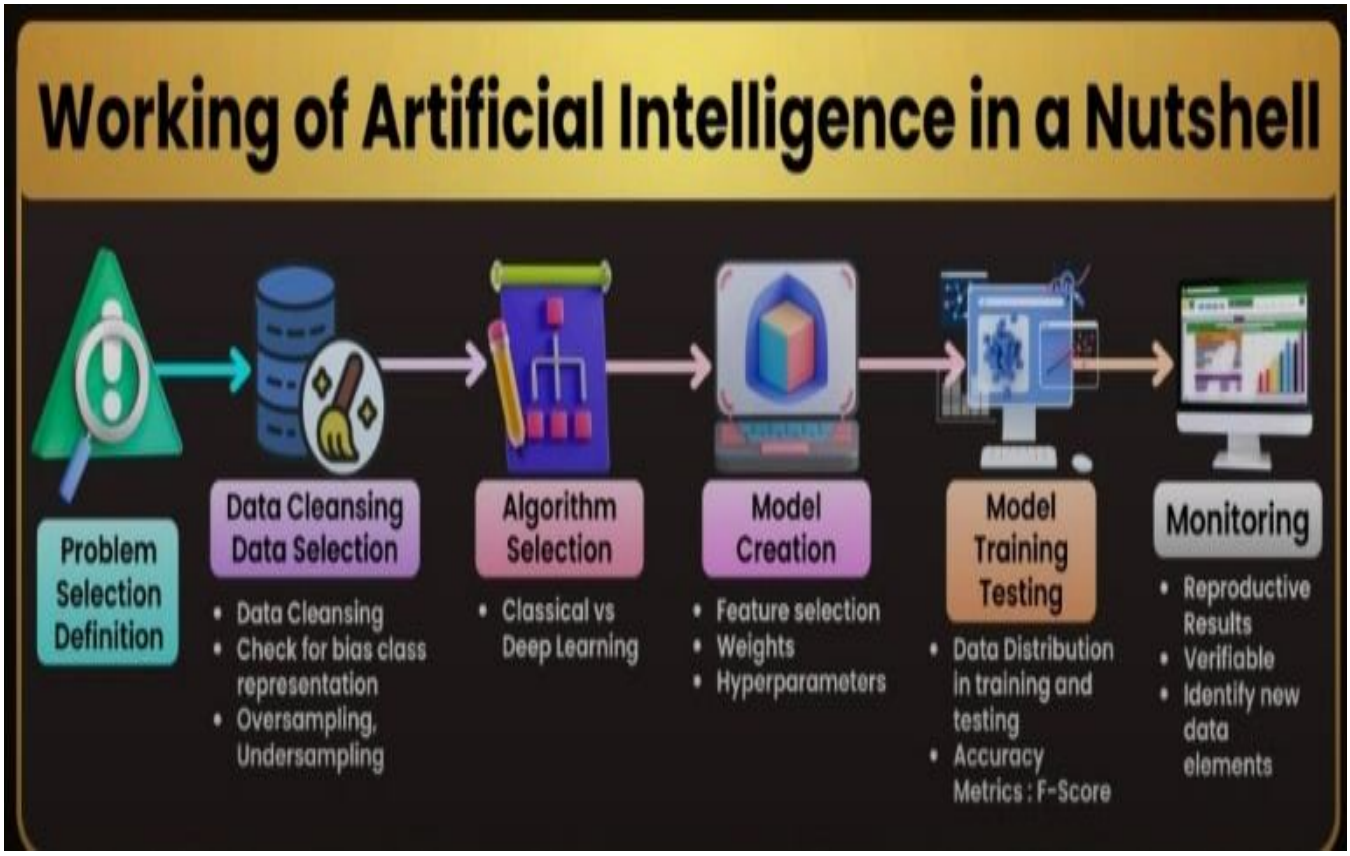


Fig. 4. Artificial Intelleigence (AI) in Nutshell

Fig. 5 explains clearly how Artificial Intelligence (AI) works as per the proposed apt process. AI process starts with defining the problem and ends with ethical consideration.

The apt process defines the following phases to solve the AI problem step by step.

A. Defining the Problem

1) Recognize and define the problem / task and

2) Specify the desired outcome and establish performance metrics.

B. Data collection and preparation:

1) Curate relevant data

2) Clean, preprocess, and annotate the data and

3) Divide data into training, validation, and test sets.

C. Selecting models and algorithms:

1) Choose a suitable AI technique

2) Develop and appropriate algorithm or model and

3) Configure model parameters accordingly.

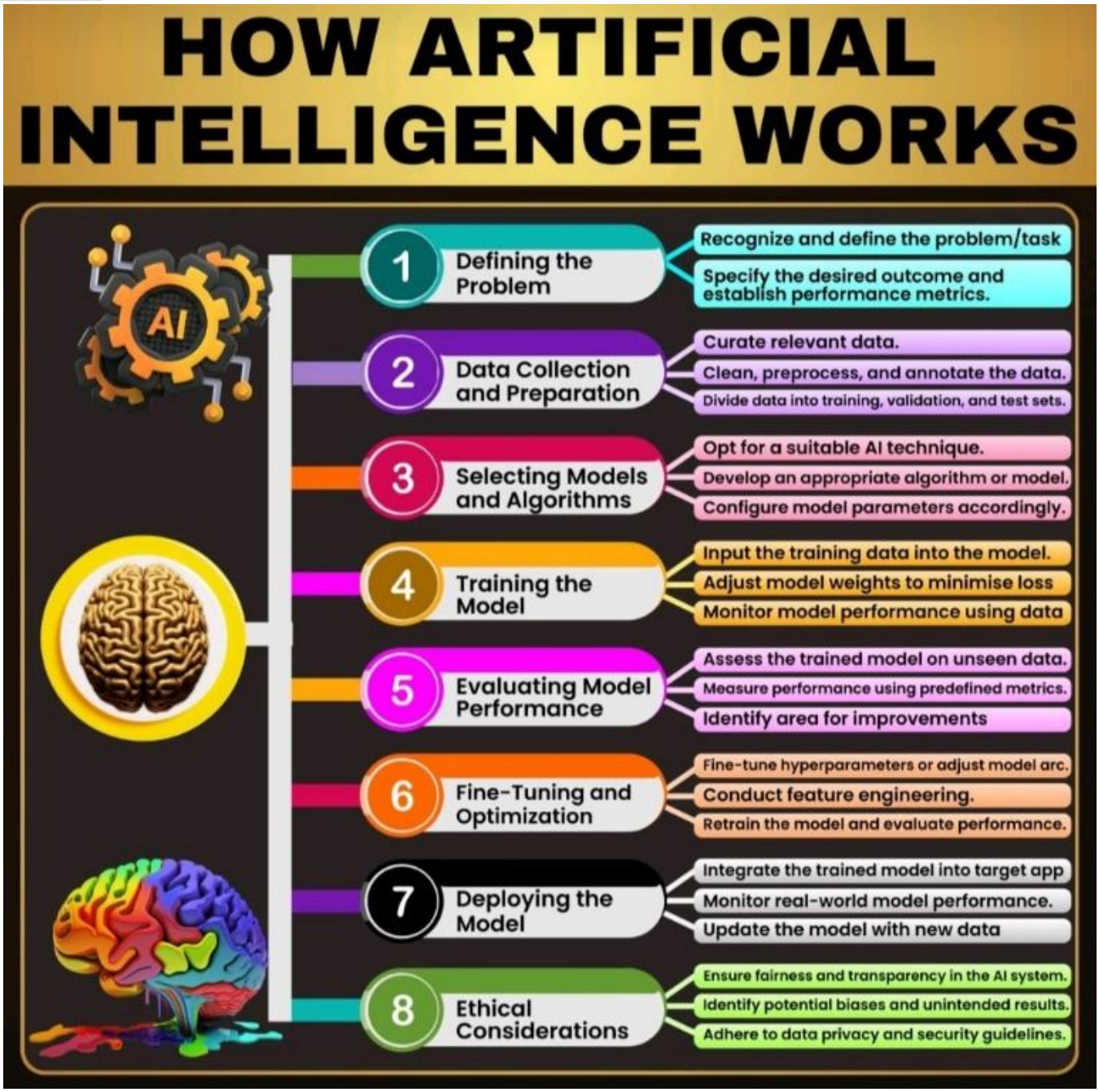


Fig. 5. Proposed apt process for Artificial Intelligence (AI)

D. Training the model:

1) Input the training data into the model
2) Adjust the model weights to minimize the loss and
3) Monitor model performance using data.

E. Evaluating the model performance:

1) Access the trained model on unseen data
2) Measure performance using predefined metrics and
3) Identify area for improvements.

F. Fine-tuning and optimization:

1) Fine-tune hyperparameters or adjust model architecture
2) Conduct feature engineering and
3) Retain the model and evaluate performance.

G. Deploying the model:

1) Integrate the trained model into target app
2) Monitor real-world model performance and
3) Update the model with new data

H. Ethical considerations:

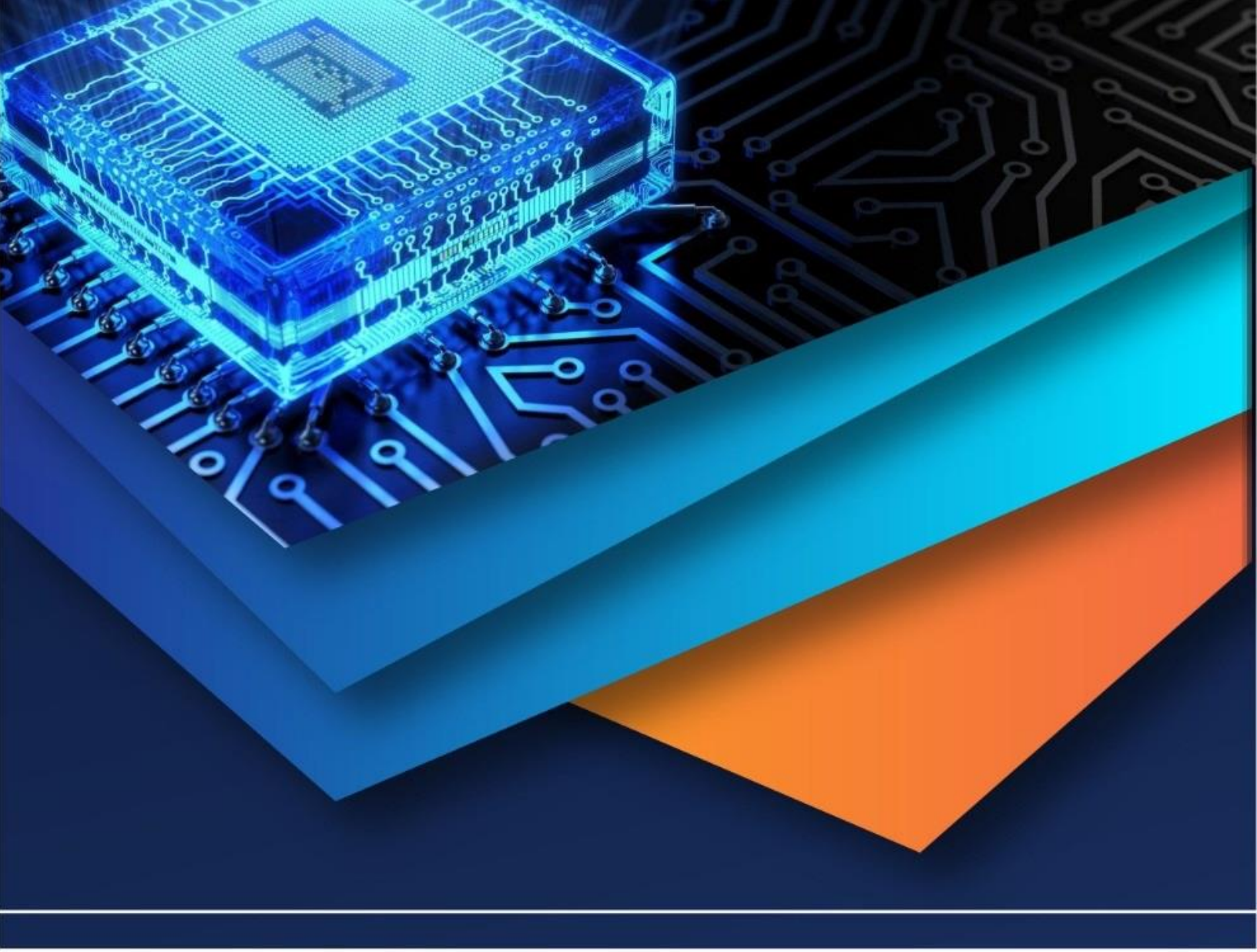
1) Ensure fairness and transparency in the AI system
2) Identify potential biases and unintended results and
3) Adhere to data privacy and security guidelines.

V. CONCLUSION

The paper proposes apt processes for Machine Learning and Artificial Intelligence. To solve the real-world problems with Machine Learning (ML) and Artificial Intelligence (AI) a definite process is proposed and which is useful while providing operative solutions. Just having or selecting great algorithms is not sufficient and does not yield solutions to the problem. Machine Learning and Artificial Intelligence required an apt process to solve a problem. Machine Learning process starts with defining the data and ends with model. An apt Machine Learning process proposed in this paper includes defining the problem, collecting the data, preparing the data, splitting the data for training and testing, algorithm selection, training the algorithm, evaluation on test data, parameter tuning, and start using the model. An apt Artificial Intelligence process proposed in this paper includes defining the problem, data collection and preparation, selecting models and algorithms, training the model, evaluating model performance, fine-tuning, and optimization, deploying the model, and ethical considerations. The proposed apt process for Machine Learning and Artificial Intelligence with step-by-step evaluation solves the real-world problems effectively for all domains.

REFERENCES

[1] Fahle, Simon, Christopher Prinz, and Bernd Kuhlenkötter. "Systematic review on machine learning (ML) methods for manufacturing processes–Identifying artificial intelligence (AI) methods for field application." Procedia CIRP 93 (2020): 413-418.
[2] S. Amershi *et al.*, "Software Engineering for Machine Learning: A Case Study," *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, Montreal, QC, Canada, 2019, pp. 291-300, doi: 10.1109/ICSE-SEIP.2019.00042.
[3] M. N. Islam, T. T. Inan, S. Rafi, S. S. Akter, I. H. Sarker and A. K. M. N. Islam, "A Systematic Review on the Use of AI and ML for Fighting the COVID-19 Pandemic," in *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 3, pp. 258-270, Dec. 2020, doi: 10.1109/TAI.2021.3062771.
[4] Bosch, Jan, Helena Holmström Olsson, and Ivica Crnkovic. "Engineering ai systems: A research agenda." Artificial Intelligence Paradigms for Smart Cyber-Physical Systems (2021): 1-19.
[5] Handelman, Guy S., Hong Kuan Kok, Ronil V. Chandra, Amir H. Razavi, Shiwei Huang, Mark Brooks, Michael J. Lee, and Hamed Asadi. "Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods." American Journal of Roentgenology 212, no. 1 (2019): 38-43.
[6] Polyzotis, Neoklis, and Matei Zaharia. "What can data-centric ai learn from data and ml engineering?." arXiv preprint arXiv:2112.06439 (2021).
[7] Kiron, David, and Michael Schrage. "Strategy for and with AI." MIT Sloan Management Review 60, no. 4 (2019).
[8] Nuseir, Mohammed T., Barween H. Al Kurdi, Muhammad T. Alshurideh, and Haitham M. Alzoubi. "Gender discrimination at workplace: Do artificial intelligence (AI) and machine learning (ML) have opinions about it." In The international conference on artificial intelligence and computer vision, pp. 301-316. Cham: Springer International Publishing, 2021.
[9] Johnson, Sandra LJ. "AI, machine learning, and ethics in health care." Journal of Legal Medicine 39, no. 4 (2019): 427-441.
[10] Jiang, Fei, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. "Artificial intelligence in healthcare: past, present and future." Stroke and vascular neurology 2, no. 4 (2017).



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)