# RESEARCH ON THE USE OF CLOUD PLATFORMS FOR TRAINING AND DEPLOYING MACHINE LEARNING MODELS AND AI SOLUTIONS

**Damodarrao Thakkalapelli**

Vice President
Bank of America
4808 Loft Ln PLano Tx 75093

***Abstract*- The machine learning model and AI solution help to train and deploy the cloud-based system which will help to analyze the machine learning model after all the analysis is stored all the data and the prediction model are in a cloud-based system with the help of API framework. In the research machine learning and Artificial networks help to do the analysis of the stored data volume in the cloud process the power and use the machine learning algorithm to find smart results and accuracy. The author believed in various cloud sets like Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, and IBM Cloud. They surveyed all the stages in light of variables like valuing, adaptability, accessible administrations, and convenience. They utilized a few Python libraries and instruments to play out the machine learning model and AI solution to train and deploy the cloud-based system**.

**Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 6, Nove-Dec 2023**



## Introduction

Cloud stages are vital in the turn of events and sending of AI models and computer-based intelligence arrangements. These stages give a universal and flexible environment that authorizes the associations to prepare and convey models productively. By using machine learning and artificial intelligence for utilizing the cloud assets, organizations can get to a solid computational foundation without the requirement for huge on-premises thoughts. Cloud stages offer broad computational assets in the trial stage, considering equal handling and sped-up model preparation. This speeds up the preparation cycle, empowering information researchers to repeat and analyze all the more quickly. For sending, cloud stages offer a strong basis and administrations to have models. They give adaptable answers for handling differing responsibilities, guaranteeing reliable execution significantly under appeal. Additionally, cloud platforms offer seamless pipelines for model deployment, facilitating automated and effective workflows and advanced networks associated with the data, and it will help to apply Deep learning to find better results.

## Use of Machine Learning and AI

Using machine learning and Artificial network help to grow the stored data volume in the cloud process the power and apply the machine learning algorithm to find smart result and accuracy. By using algorithms, the cloud-based algorithm is rapidly concentrated on the system and the aspects to sustain the machine learning on giant data sets. All the cloud providers also focused on Artificial intelligence for better analysis of cloud platforms for better and faster work provider platforms [4]. Artificial intelligence provides the application that developers can use the cloud computing infrastructure to train the machine learning models and it will help further computer services such as server-less computing and batch processing and for automatic machine learning work pressure. Cloud stages offer an option by giving virtualized conditions that can powerfully designate assets in light of responsibility necessities.

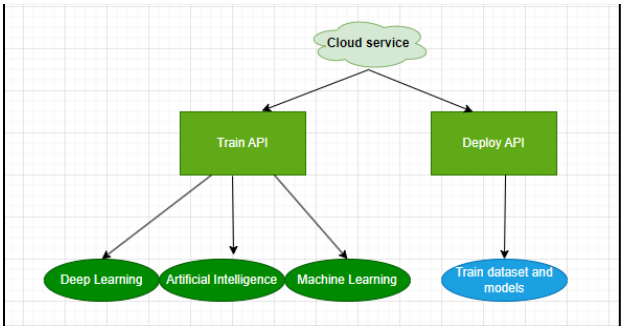


**Figure 1: Machine learning in the cloud-based system**
(Source: Self-created in draw.io)

The machine learning algorithm and artificial intelligence developed upon progress in distributed computing innovation, which has seen outstanding development lately and here they use machine learning models like KNN and logistic regression.

**KNN Algorithm-** KNN is known as K-Nearest Neighbours is an essential classification algorithm of Machine Learning. KNN algorithm only applies to the supervised type of data after train and testing the algorithm applied to train the cloud data deploying the data in the cloud best system to get more reliable and accurate results. In the algorithm first, it needs to find the neighbor and calculate the Euclidean distance Square root( $(X^2\text{-}X^1)^2+(Y^2\text{-}Y^1)^2$) by using this formula they used the data point distance, and then the model is prepared and then it needs to train the data.
**Linear Regression-** It is also a classification algorithm that depends upon the data variable and it was also done by using by classifying the model and applying a supervised learning approach and the algorithm applies the straight line formula to find the distance of data and analysis of data.

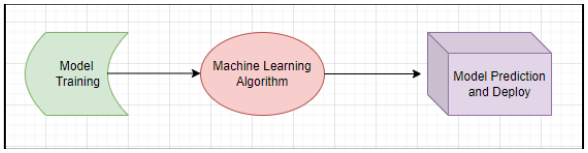


**Figure 2: Workflow of Machine Learning Algorithm**
(Source: Self-created in draw.io)

Accessing powerful computing resources has been democratized by the media of Platform as a Service and Infrastructure as a Service offering, leveling the playing field for businesses of all sizes. Also, the appearance of specific equipment, like GPUs and TPUs, has reformed the speed and productivity of model preparation [5]. The cloud-based application suggests the speed of the organization as well as the adaptability and determines the quality of man-made intelligence arrangements. The cloud-based AI and simulated intelligence sending tend to a basic gap in the field and give an association to the improvement of universal, open, and cost-effective deals that can drive growth across different enterprises.

**Materials and Methods**
Here is a Materials and Strategies segment for this exploration on the utilization of cloud stages for preparing and conveying AI models and simulated intelligence arrangements utilizing Python. To direct this exploration, that previously assembled significant information and datasets from different sources that are ordinarily utilized in AI and artificial intelligence projects. The information sources included openly accessible datasets, exclusive datasets, and mimicked information [6]. That considered various cloud stages for this research, including Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, and IBM Cloud. They surveyed these stages in light of variables like valuing, adaptability, accessible administrations, and convenience [3]. They utilized a few Python libraries and instruments to play out the research including yet not restricted. a broad assortment of information was collected to guarantee a different dataset delegate of true situations. Information was obtained from different sources, including open information vaults, industry-explicit data sets, and custom information age.

**Scikit-Learn**- for the creation and evaluation of machine learning models.
**TensorFlow and PyTorch**- For deep learning model turn of events.
**Jupyter Notebooks**- For intelligent code advancement and documentation.
**Pandas, and NumPy**- For information control and preprocessing.
**Matplotlib and Seaborn**- For information representation.

To choose the best platform for our research's specific tasks, each platform was thoroughly evaluated [7]. Python was selected as the primary programming language due to its adaptability and extensive library ecosystem. For this utilization this is a scope of libraries and instruments for information research, model turn of events, and cloud joining, including TensorFlow, PyTorch, Keras, Scikit-Learn, Boto3 (for AWS), and SDK for Python [2]. This research preprocessed the data by performing operations like data cleaning, feature engineering, and handling missing values before training any machine learning models. This step guaranteed that the information was appropriate for model preparation. Using Python, one can create a variety of deep learning and machine learning models.

**Models of Classification-** For undertakings like feeling investigation and picture grouping.
**Neural Networks-** For complex profound learning undertakings.

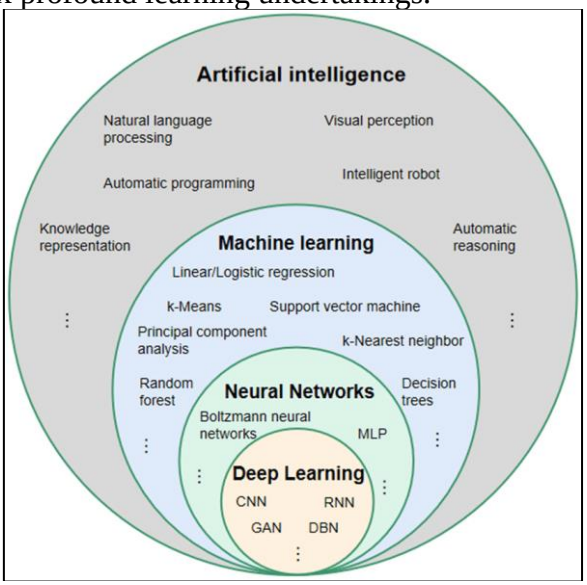


**Figure 3: Relationship between neural network, machine learning, and AI**
(Source: https://www.researchgate.net)

The infrastructure and resources required for model training and deployment were installed on each cloud platform. They arranged virtual machines, GPU cases, and the capacity to oblige the necessities. Utilizing the chosen cloud stages, which are prepared through AI and deep learning models [1]. They followed preparing measurements like exactness, misfortune, and assembly time for assessment. Money-saving advantage research was performed to decide the most expense-proficient stage for explicit use cases [8]. Results were envisioned utilizing heatmaps, bar graphs, and line plots to work with clear understanding.
After model preparation, they sent the models to the cloud stages. Then likewise directed A/B testing and contrasted results with the survey and their genuine relevance [10]. One can examine the cloud platforms' performance metrics and the data gathered. In the analysis, this took into account costs, deployment ease, and scalability. Suitable graphs and charts were used to present the results visually. When working with datasets and cloud platforms, they made sure that this research followed ethical guidelines and complied with privacy and data protection regulations.
Data preprocessing was a multi-stage procedure that included removing outliers and anomalies from the data, scaling and transforming features, and dealing with missing values. Cautious thought was given to keep up with information respectability and limit inclination [9]. To answer a variety of research questions, several machine learning and deep learning models were created. These models enveloped straight and non-direct

calculations, convolutional and repetitive brain organizations, and transformer-based designs. Hyperparameter tuning was directed utilizing procedures, for example, framework search and Bayesian enhancement.

**Analysis**

The cloud-based solution helps to build the runtime analysis service for deep learning and machine learning in the cloud has both system-related challenges and algorithmic challenges.

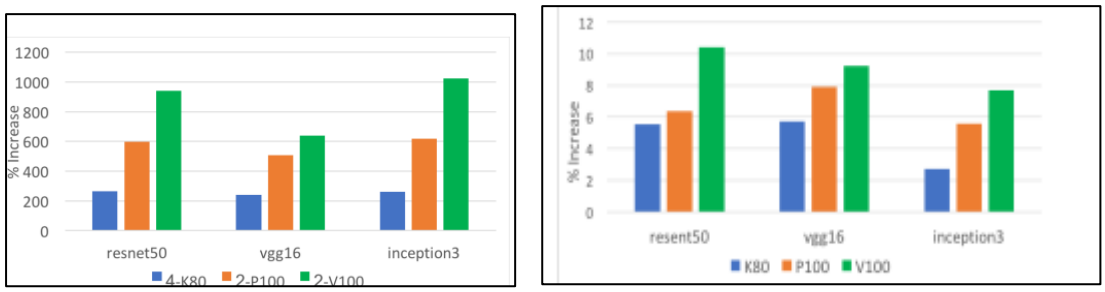


**Figure 4: Training time speed-up of DL models on different GPUs compared to a single**
(Source: https://www.researchgate.net.pdf)

In the above snip of the Training time speed-up of deep learning model on different GPUs compared to a single mainly analyzes the run time speed and concerns the deep learning model and the GPU runtime speed is increasingly high with the complicated way of runtime [11]. The batch altitude influences the runtime of the GPUs because it helps to control the quantity of data processed per total cycle, and hence the compute-to-communication ratio. The snips also show the scaling of throughput when it is growing the set size from 32 to 64 for types of GPUs. The goal of the design is to understand the service to work with cloud platforms supporting diverse types of hardware resources and Deep learning frameworks, The need for a modeling approach helps to catch the connection between the multiple content and runtime, usually, the inter-dependent aspects are connected to a software model and resource.

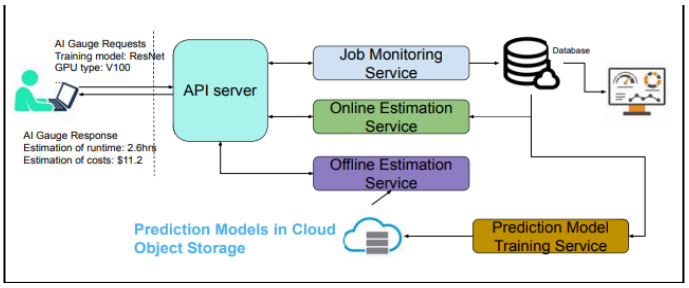


**Figure 5: AI Gauge Service architecture**
(Source: https://www.researchgate.net.pdf )

In the above snip of the photo, the current cloud-based deep learning offering space is highly competitive primary providers offer their customers a variety and different tools and frameworks for their deep learning investigations [12]. The AI gauge works with the training model and the runtime to the API server and Job monitoring services. Then it goes to the specific dataset and trains the prediction model according to the service and at least it goes to the cloud objects storage of different prediction models.

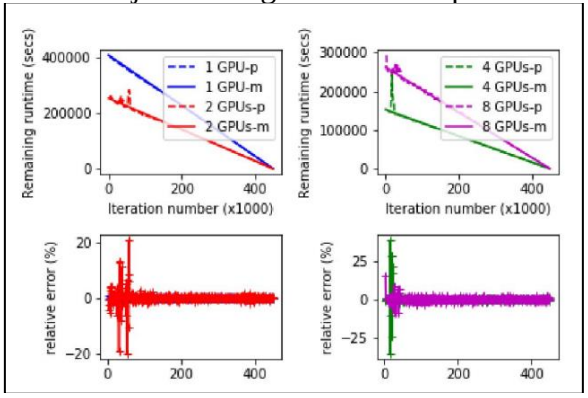

**Figure 6: Comparison of measured and estimated remaining runtime**
(Source: https://www.researchgate.net.pdf )

The above figure snip reaches the remaining runtime part and find various type of graphs and they find various type of runtime in the part. They find the different runtime for the machine learning model.
"y˜ = β0 + 4 ∑ i=1 βiIi + β5W + β6 B T + 1 + β7 B W × F "
After applying all the models they find this equation. Here, in the equation l is the variable and y is the estimated time, w is the number of work how many times the algorithm did and here in the graph, it shows the decreases in run time.

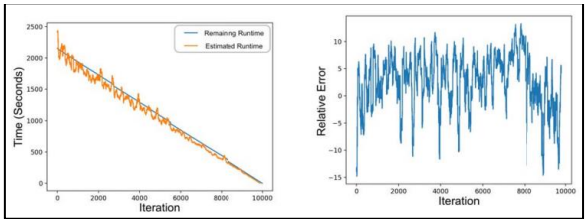


**Figure 7: Estimated remaining runtime and error variation**
(Source: https://www.researchgate.net.pdf)

The above snip shows the calculated renaming runtime of the model and the runtime iteration.
"1/ |T| ∑ (x,y)∈T |y − φ(x)| y "
In this equation y is the dependent variable, and φ(x) is the model calculated value and T is the set of tests of how many types it occurs.

**Discussion**
The cloud-based analysis of machine learning in the cloud-based system helps to find the visual representation of the speed of train time concerning the deep learning model dataset and here the analysis is mainly about the run time speed and different types of GPU. GPU helps to control the quantity of the processed data depending upon per cycle count and it helps to find the communication ratio. From the analysis the scaling of throughput when it is
growing and it works with 32 and 64 sizes of data in GPUs. In the analysis part they work with the training time speed concerning the deep learning model and do the design to understand all the work and service of the cloud supporting service work and the types of software resources and Deep learning frameworks. Also, from the analysis work with the AI gauge training and the response of the AI gauge which first goes to the API server and provides various types of services with the help of the API server mainly the server did job monitoring service and estimation service and those services are first store into the database and it stores all the data. After storing all the data in the database it helps to do the prediction model training service and at the last stage, all the stored data go to store all the prediction models in cloud object storage.

**Conclusion**
The cloud-based analysis of machine learning models stages for preparing and conveying AI models and simulated intelligence arrangements. The discoveries exhibit that a cloud-based foundation offers critical benefits regarding versatility, execution, and cost-viability contrasted with conventional on-premises arrangements. Particular equipment further enhances these advantages. Additionally, cloud-native deployment mechanisms simplify the procedure and guarantee consistent and dependable performance across a variety of workloads. These bits of knowledge make ready for a change in outlook in artificial intelligence improvement, democratizing admittance to strong processing assets and empowering development across different ventures. Embracing cloud stages is basic for associations looking to outfit the maximum capacity of AI and artificial intelligence advancements and it is shown in the visual representation plot of train time speed of deep learning and the storage procedure of machine learning model into cloud database

**REFERENCES:**

**Journals**


1. Hummer, W., Muthusamy, V., Rausch, T., Dube, P., El Maghraoui, K., Murthi, A. and Oum, P., 2019, June. Modelops: Cloud-based lifecycle management for reliable and trusted AI. In *2019 IEEE International Conference on Cloud Engineering (IC2E)* (pp. 113-120). IEEE.
2. Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B. and Zimmermann, T., 2019, May. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)* (pp. 291-300). IEEE.
3. Garg, S., Pundir, P., Rathee, G., Gupta, P.K., Garg, S. and Ahlawat, S., 2021, December. On continuous integration/continuous delivery for automated deployment of machine learning models using mlops. In *2021 IEEE fourth international conference on artificial intelligence and knowledge engineering (AIKE)* (pp. 25-28). IEEE.
4. Dube, P., Suk, T. and Wang, C., 2019, October. AI gauge: Runtime estimation for deep learning in the cloud. In *2019 31st International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)* (pp. 160-167). IEEE.
5. Murshed, M.S., Murphy, C., Hou, D., Khan, N., Ananthanarayanan, G. and Hussain, F., 2021. Machine learning at the network edge: A survey. *ACM Computing Surveys (CSUR)*, *54*(8), pp.1-37.
6. Wang, X., Han, Y., Leung, V.C., Niyato, D., Yan, X. and Chen, X., 2020. Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *22*(2), pp.869-904.
7. García, Á.L., De Lucas, J.M., Antonacci, M., Zu Castell, W., David, M., Hardt, M., Iglesias, L.L., Moltó, G., Plociennik, M., Tran, V. and Alic, A.S., 2020. A cloud-based framework for machine learning workloads and applications. *IEEE access*, *8*, pp.18681-18692.
8. Philipp, R., Mladenow, A., Strauss, C. and Völz, A., 2020, November. Machine learning as a service: Challenges in research and applications. In *Proceedings of the 22nd International Conference on Information Integration and Web-based Applications & Services* (pp. 396-406).
9. Battina, D.S., 2019. An intelligent devops platform research and design based on machine learning. *training*, *6*.
10. Christidis, A., Davies, R. and Moschoyiannis, S., 2019, November. Serving machine learning workloads in resource constrained environments: A serverless deployment example. In *2019 IEEE 12th Conference on Service-Oriented Computing and Applications (SOCA)* (pp. 55-63). IEEE.
11. Wan, C., Liu, S., Hoffmann, H., Maire, M. and Lu, S., 2021, May. Are machine learning cloud apps used correctly? In *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)* (pp. 125-137). IEEE.
12. Merenda, M., Porcaro, C. and Iero, D., 2020. Edge machine learning for ai-enabled IoT devices: A review. *Sensors*, *20*(9), p.2533.