Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

An Investigation Of Automating Datacenter Operations Through ML Algorithm

[1]Ujjal Aloke Sarkar , [2] Bhupender Kumar Singh

[1]Ph.D Research Scholar, Dept. of Computer Science and Engg., Shri Venkateshwara University , Gajraula (UP)

[2]Associate Professor, IIMT University, Meerut (UP)

Abstract: - Machine applications bring new workflow and challenges in the design of enterprise data. One of the key challenges relates to the data and storage solutions needed to store, manage, and deploy data services. Today's smart applications require infrastructure that is vastly different from traditional analytics operations, and decisions to build organizational data will have a significant impact on the success of data-driven projects. The focus in data center is the data itself. There are various types of data like any unresolved fact, amount, text, sound, or image that can be defined and analyzed. It is the most important part of Data Analytics, Machine Learning. A combination of import information, experiences, learning and understanding. Effects of data optimized through certain machine learning algorithm on the awareness or mindset of an individual or organization. This thesis explores the two different techniques for the study of data center through machine learning. This has two different mechanisms for the study one is for the security of data in transmission second is for the data processing. This overall simulation which has focused on the data center operation executed through MATLAB. With this simulation tools this thesis explores the two major study. This study only includes main and recent machine learning algorithm which executed on MATLAB. Result in graphical and analytical form shows the data center operations in real time works.

Keywords: - Data center, Data Analytics, Machine Learning, Data center operation

I. INTRODUCTUON

Good performance and high availability are the primary goals of companies running Web applications that we rely on in our daily lives, such as Google or Facebook. Once Web applications are deployed to a production environment, the task of datacenter operators is to constantly monitor the applications, notice possible performance issues, and quickly resolve them. Because application downtime could potentially cost millions of dollars, the operators are often required to respond to an alert within 10 or 15 minutes, even at night or on the weekends. Due to increasing complexity and large-scale deployment of popular Web applications, the job of datacenter operators is becoming more difficult. For example, Web applications are often deployed in multiple, globally-distributed datacenters for improved performance and redundancy. Typical datacenters run tens of thousands of commodity servers that frequently fail; during the first year of a typical 2,000-node cluster, Google experiences 20 rack failures, 1000 machine failures, thousands of disk failures as well as many other types of failures [Dean, 2009]. The functionality of Web applications is often provided by tens or hundreds of independent software services that are constantly evolving. For example, eBay deploys 300 features per quarter and adds 100,000 new lines of code every two weeks [Shoup, 2007]. Moreover, requests of different applications share the same hardware or network equipment. Artificial Intelligence (AI) and Machine Learning (ML) technologies are a proven way for data center operators to maximize uptime, optimize energy usage, quickly detect potential risks and defend against cyber-attacks. So, it's no surprise that 83 percent of organizations have increased their AI/ML budgets year-on-year, according to Alorithmia's "2021 Enterprise Trends in Machine Learning." For example, major hyper scalers have developed in-house AI to support use cases such as cooling. But smaller operators can achieve AI/ML benefits too, by leveraging AI-as-a-Service on cloud platforms.

Embedded AI

Data center vendors are also steadily making it easier for their customers to begin using AI/ML by embedding the technology in their products. One example is specialized silicon designed to perform complex mathematical and computational tasks in a more efficient way. Most AI use cases today are very narrow, so these AI chips can be trained for a specific task such as pattern recognition, natural language processing, network security, robotics and automation. AI is maturing, which means its capabilities are growing at the same time it rides down the cost curve. Those two trends will enable data center vendors to embed AI/ML into more of their products. For example, RISC-V and other open-source technologies are lowering the barriers to purpose-built "building blocks" that can focus on efficiency, performance and scalability like never before. That in turn will drive even more adoption and use cases, including among smaller data center operators that currently consider AI/ML too expensive to implement widely or at all. The AI/ML can be applied to the data center's mechanical and electrical equipment to enable actionable insights and automation, saving money for the operator. This requires integrating traditional physics-based modelling approaches with state-of-the-art ML techniques using data from Internet of Things (IoT) sensors. ML and physics-based modelling both have their strengths. Combining them leverages the best of both worlds to solve complex data center issues involving mechanical and electrical equipment. With 5G and related industry 4.0 use cases there is a steep increase in demand for 'anywhere, anytime' access to applications and services such as autonomous vehicles, smart cities, advanced manufacturing, AR/VR gaming etc. Latency is no longer tolerable. As a result, edge data centers are taking center stage as are multi-access edge compute (MEC) capabilities. With compact, inexpensive and powerful hardware in edge data centers, it is now possible to run AI/ML workloads close to the user where data is generated, and get real-time insights and experiences, delivered by highly responsive and contextually aware apps.

Machine Learning Is Helping Data Centers

Deep Machine Learning is used to examine huge datasets to identify and find out patterns and outline the human interface on the prediction based out of the data. Data centers are equipped with sensors to filter out and provide historical data. Many research centers apply (ML) and (AI) on their historical data to improve efficiency and productivity.

Developed Engineering

(ML) and (AI) technology is used in data centers by engineers and researchers to design and implement recommendation engines. Ever imagined when the Air Conditioner in our room meets the specified temperature, it automatically cuts off the cooling. This is (ML) that is working in the background as programmed by you through the remote. When the room temperature reaches five degrees above the targeted temperature, it automatically turns ON the cooling as the programmed data instructs it to do so. This was a small example however we can understand that such sensors are responsible to collect data and send it to data centers. (AI) and (ML) uses this data to analyze and identify the pattern and process instructions based on our programming. The (AI) is also responsible for energy efficiency, power management, cost reduction, accuracy, eliminate human interface and meet all tasks by machine learning and artificial intelligence.

Data Security

Ransomware is the biggest cybersecurity threat that requires immediate attention. Every organization is working day and night to ensure military-grade security on their data centers. (ML) and (AI) are continuously a supporting pillar in constantly improving data security. AI has a significant impact on the way data centers are keeping the hackers away from any such sort of attack. (AI) process large amount of data that a human can't process with his eyes and brain and in no-time, (AI) can identify and eliminate any suspicious or possible threat.

Adding Competitive Advantages to Data Science

Every company needs reliability and advanced data science for swift functionalities and a confident business environment. This all is made possible with the evolution of (AI) and (ML) that revolutionary evolved high-tech updates and demonstrated.

Advancement.

Nowadays, companies choose using data centers instead of local servers as data centers are more secured, reliable and loaded with high-tech advancements to process business needs without any lag. A few of the (AI) based data centers keep their data on cloud-based servers. These servers are easy to access, secured and successful in business processing and automation. Data science has evolved to such a state, that it can meet expertise while keeping deadlines intact.

Reducing Need for Data Scientists

AI has become so advanced that now we assume, it is going to harm the employment of thousands of data scientists. These super-computers can use (AI) in such a way that these can furnish the tasks of hundreds of employees just by itself. AI is going to reduce the need of data scientists as its advanced algorithms can process all the mathematical calculations without any break. These data centers can store, process, analyze and simulate data of various clients without any human interface. There is nobody required to maintain cloud-based data centers as they work on auto-pilot mode. The hassle-free transformation of data centers and data science is the quest of this era.

Use AI/ML for new construction

Operators should make AI/ML a key part of their planning and construction process, such as with building information modelling (BIM) and building performance simulation (BPS) tools. This advice applies to retrofit projects, such as enabling predictive maintenance at an existing facility. To ensure a successful retrofit:

- **Develop a retrofit strategy**
  Identify the business objectives of retrofitting, including which machines in the installed base will be "brought online" and the potential sales of digital services and associated costs.

- **Develop a data strategy**
  For companies working in a legacy manufacturing environment, a significant concern is lack of real-time visibility into operations. Accessing data in legacy systems is challenging. Even if a legacy system can generate data, the reports often arrive days or weeks later sometimes too late to do anything about a problem.

- **Choose the right set of hardware and software solutions**
  These should facilitate the connection of assets regardless of their type, brand, age, protocol or communication standard.

- **Put security at the forefront**
  Ensure that the retrofit solution uses encryption and development approaches such as security by design and Trusted Execution Environment (TEE) procedures.

II. BACKGROUND

Hazelwood, et al. (2018), Machine learning sits at the core of many essential products and services at Facebook. This paper describes the hardware and software infrastructure that supports machine learning at global scale. Facebook's machine learning workloads are extremely diverse: services require many different types of models in practice. This diversity has implications at all layers in the system stack. In addition, a sizable fraction of all data stored at Facebook flows through machine learning pipelines, presenting significant challenges in delivering data to high-performance distributed training flows. Computational requirements are also intense, leveraging both GPU and CPU platforms for training and abundant CPU capacity for real-time inference. Addressing these and other emerging challenges continues to require diverse efforts that span machine learning algorithms, software, and hardware design.

Chen et al. (2018), traffic optimizations in datacenters are difficult online decision-making problems. Previously, they are done with heuristics relying on operators' understanding of the workload and environment. Designing and implementing proper TO algorithms thus take at least weeks. Encouraged by recent successes in applying deep reinforcement learning (DRL) techniques to solve complex online control problems, we study if DRL can be used for automatic TO without human-intervention. However, our experiments show that the latency of current DRL systems cannot handle flow-level TO at the scale of current datacenters, because short flows (which constitute the majority of traffic) are usually gone before decisions can be made.

Todd et al. (2021), HPC data centers such as the one at NREL's ESIF will increasingly need to rely on automation to keep pace with exascale growth in compute capability and to manage and optimize the data center environment and facility resources. Artificial intelligence (AI) and machine learning (ML) approaches provide the means to improve HPC data center efficiency (energy, operational, and managerial efficiency) and resiliency by learning historical trends and training models to operate on real-time data collected from both IT and facilities sources. The goal of coupled improvement of data center resiliency and energy efficiency through automated data collection and AI has led to a multi-year, multi-staged collaboration between NREL and Hewlett-Packard Enterprise's Advanced Technology Group, referred to as Artificial Intelligence for Data Center Operations (AIOps). The extended efforts within the AIOps project include a common goal of building capabilities for an advanced smart facility and demonstration of data collection and AI modeling techniques in the ESIF data center.
Berral et al. (2010), As energy-related costs have become a major economical factor for IT infrastructures and data-centers, companies and the research community are being challenged to find better and more efficient power-aware resource management strategies. There is a growing interest in "Green" IT and there is still a big gap in this area to be covered.
In order to obtain an energy-efficient data center, we propose a framework that provides an intelligent consolidation methodology using different techniques such as turning on/off machines, power-aware consolidation algorithms, and machine learning techniques to deal with uncertain information while maximizing performance. For the machine learning approach, we use models learned from previous system behaviors in order to predict power consumption levels, CPU loads, and SLA timings, and improve scheduling decisions. Our framework is vertical, because it considers from watt consumption to workload features, and cross-disciplinary, as it uses a wide variety of techniques. We evaluate these techniques with a framework that covers the whole control cycle of a real scenario, using a simulation with representative heterogeneous workloads, and we measure the quality of the results according to a set of metrics focused toward our goals, besides traditional policies. The results obtained indicate that our approach is close to the optimal placement and behaves better when the level of uncertainty increases.

Mirhoseini Nejad et al. (2021), Two key contributors to the energy expenditure in data centers are information technology (IT) equipment and cooling infrastructures. The standard practice of data centers lacks a tight correlation between these two entities, resulting in considerable power wastage. Considering the cooling cost of different locations inside a data center (cooling heterogeneity) and various cooling capabilities of servers (server heterogeneity) has significant potential for saving power, yet has not been studied thoroughly in the literature. There is a necessity for state-of-the-art approaches to integrate the control of IT and cooling units. Moreover, the literature still lacks an accurate and fast thermal model for temperature prediction inside a data center. In this paper, innovative approaches to quantify data center thermal heterogeneities are presented. Using data center thermal models, the cost of providing cold air at the front of servers can be (indirectly) calculated, and the capability of servers to be cooled is formulated. Our approach assigns jobs to locations that are efficient to cool (from the perspectives of both servers and cooling units) and tunes cooling unit parameters. The method, called holistic data center infrastructure control (HDIC), has the potential to save a considerable amount of power by exploiting synergies between the workload scheduler and operational parameters of cooling units.

Haghshenas et al. (2019), Improving the energy efficiency of data centers while guaranteeing Quality of Service (QoS), together with detecting performance variability of servers caused by either hardware or software failures, are two of the major challenges for efficient resource management of large-scale cloud infrastructures. Previous works in the area of dynamic Virtual Machine (VM) consolidation are mostly focused on addressing the energy challenge, but fall short in proposing comprehensive, scalable, and low-overhead approaches that jointly tackle energy efficiency and performance variability. Moreover, they usually assume over-simplistic power models, and fail to accurately consider all the delay and power costs associated with VM migration and host power mode transition. These assumptions are no longer valid in modern servers executing heterogeneous workloads and lead to unrealistic or inefficient results. In this paper, we propose a centralized-distributed low-overhead failure-aware dynamic VM consolidation strategy to minimize energy consumption in large-scale data centers. Our approach selects the most adequate power mode and frequency of each host during runtime using a distributed multi-agent Machine Learning (ML) based strategy, and migrates the VMs accordingly using a centralized heuristic. Our Multi-Agent machine learning-based approach for Energy efficient dynamic Consolidation (MAGNETIC) is implemented in a modified version of the Cloud Sim simulator, and considers the energy and delay overheads associated with host power mode transition and VM migration, and is evaluated using power traces collected from various workloads running in real servers and resource utilization logs from cloud data center infrastructures. Results show how our strategy reduces data center energy consumption by up to 15% compared to other works in the state-of-the-art (SoA), guaranteeing the same QoS and reducing the number of VM migrations and host power mode transitions by up to 86% and 90%, respect.

Teixeira et al. (2010), Supervisory processes are fundamental when running data center operations striving for fault resilience: any downtime can directly affect the business's income and definitely its reputation. Current monitoring tools rely on experts to configure constant thresholds on single streams, which is not appropriated for dynamic systems and insufficient to capture complex patterns. We present HOLMES, built to support data center experts to anticipate failures with a solution that combines Event Driven Architecture, Complex Event Processing and an unsupervised machine learning algorithm. Based on rules created by the users, the system continuously checks for known problems. Meanwhile, for the unknown ones, we leverage the CEP engine for aggregating and joining streams of real-time data to feed normalized input to FRAHST, our machine learning algorithm that detects anomalous patterns across multivariate numerical streams. We describe how the UI module also operates within the publish/subscribe paradigm to enhance situational awareness. The system had very well acceptance and was successfully implemented at one of the largest Internet Service Providers in South America.

Tarutani et al. (2015), To decrease the power consumption of data centers, coordinated control of air conditioners and task assignment on servers is crucial. It takes tens of minutes for changes of operational parameters of air conditioners including outlet air temperature and volume to be actually reflected in the temperature distribution in the whole data center. Proactive control of the air conditioners is therefore required according to the predicted temperature distribution, which is highly dependent on the task assignment on the servers. In this paper, we apply a machine learning technique for predicting the temperature distribution in a data center. The temperature predictor employs regression models for describing the temperature distribution as it is predicted to be several minutes in the future, with the model parameters trained using operational data monitored at the target data center. We evaluated the performance of the temperature predictor for an experimental data center, in terms of the accuracy of the regression models and the calculation times for training and prediction. The temperature distribution was predicted with an accuracy of 0.095°C. The calculation times for training and prediction were around 1,000 seconds and 10 seconds, respectively. Furthermore, the power consumption of air conditioners was decreased by roughly 30% through proactive control based on the predicting temperature distribution.

III. METHODOLOGY

3.1 General Background

The proposed work to show the data center operation using machine learning through MATlAB-2013. This will be executed through two separate buttons. One is for the understanding the machine learning behavior and second is for the data processing. The overall GUI consists two button which accumulative represent the scenario of the data center operations. As data center must be self-decisive so implementation of AI must be necessary.

Artificial Intelligence (AI) dominates today's technological discussion. AI has been a desired goal since 1956 when it was first described at the Dartmouth College workplace. The original purpose of AI was to invent machines that could mimic human learning and intelligence. Since then, this broad definition has been further refined. Today AI can say:

- Neural Networks - Complex learning and applied to high-level problems.
- Machine Learning - Reads through information gathered.
- Deep Learning - Learning about data analysis and data blocking.

Much of the recent AI revolution is based on advances in machine learning and deep learning. It has a wide, practical application that is used in a variety of industries. Self-driving cars, digital assistants, face-to-face cameras, powerful peripheral gadgets, in-depth medical diagnostic tools and much more are all direct results of the new AI. Data centers have also begun to take advantage of the AI movement. First, it helps data centers work more efficiently while reducing costs. But there are other possibilities. So one part machine learning is also very focused area of this thesis.

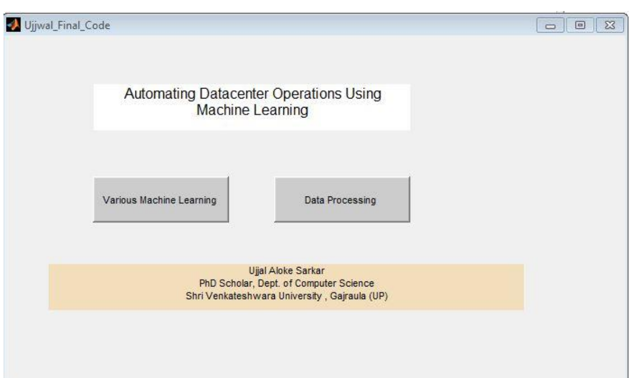


Fig. 1: It shows the GUI of the execution of the data center operation using two different procedures.

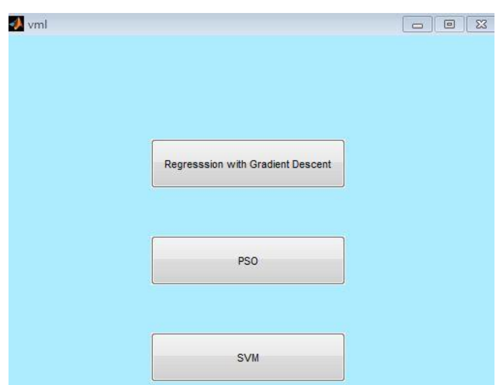


Fig. 2: this layout having three buttons (regression, PSO, SVM)

The execution of two will provide the processing of two simulative behavior that will raise the data center behavior. Machine learning is exciting in the industry these days. Organizations around the world want to integrate machine learning into their operations and new opportunities for aspiring emerging data scientists. But we have seen a huge gap between what the sector needs and what it currently offers. At the end of this page, this will understand not only what machine learning is but also its variety, its ever-expanding list of applications, the latest advances in machine learning, top experts in machine learning, among other things. The figure 2 – layout comes after this execution of first button of the main GUI. This figure having three buttons namely regression, PSO and SVM. These three buttons represent the major machine learning algorithms which is mostly used in data center operation.

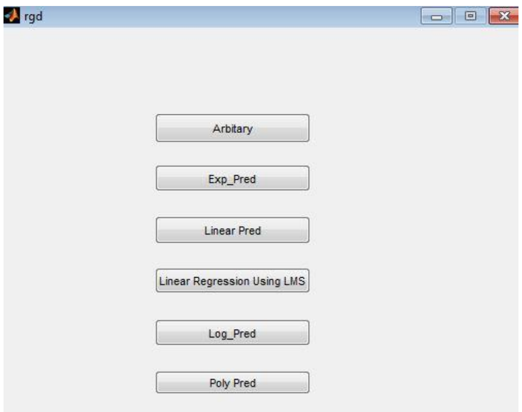


Fig. 1: This layout is for the data processing

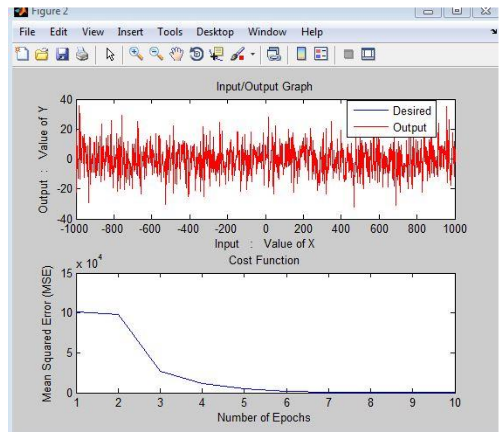


Fig. 2: This layout shows the output of the desired and the input values of data center.

This figure 3 layout come after the execution of second button. This layout consists of the six buttons with arbitrary, exp_pred, linear pred, Linear regression using LMS, log_pred and poly pred. This figure 4 demonstrate after the execution of data (arbitrary). This also present the mean square error with number of epochs. As we see that the increasing of epochs will let down the value of MSE.

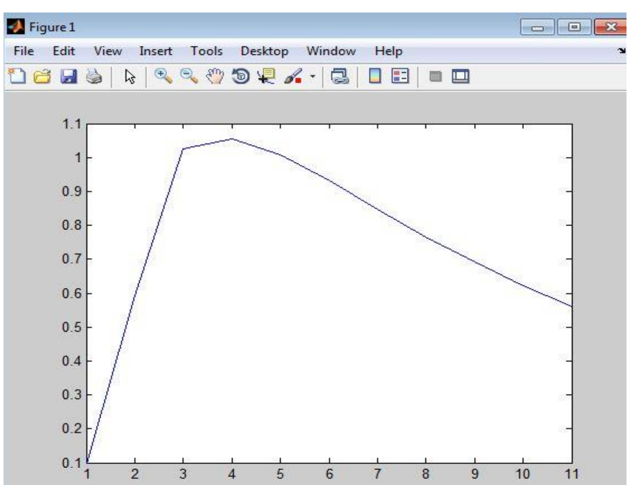


Fig. 3: This figure is simple time vs. peak graph.

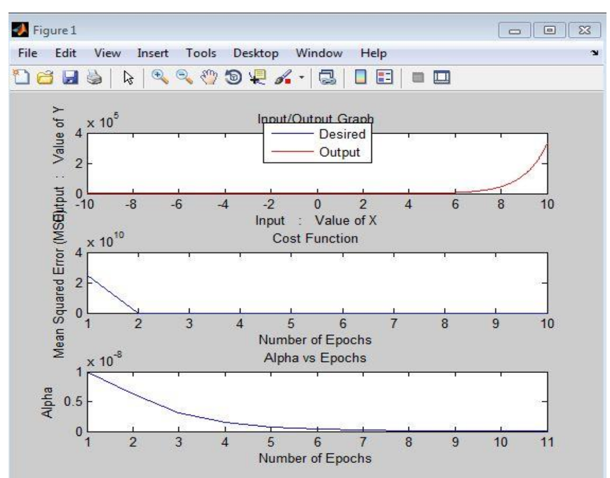


Fig. 4: shows the output vs. mean square error with alpha.

This figure 5 demonstrate after the execution of data (others). As we see that the increasing of time will let down the value of peaks.

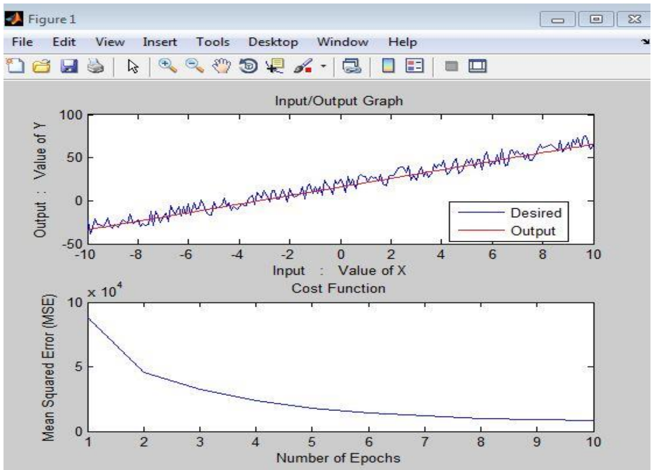


Fig. 5: figure shows the output vs. mean square error

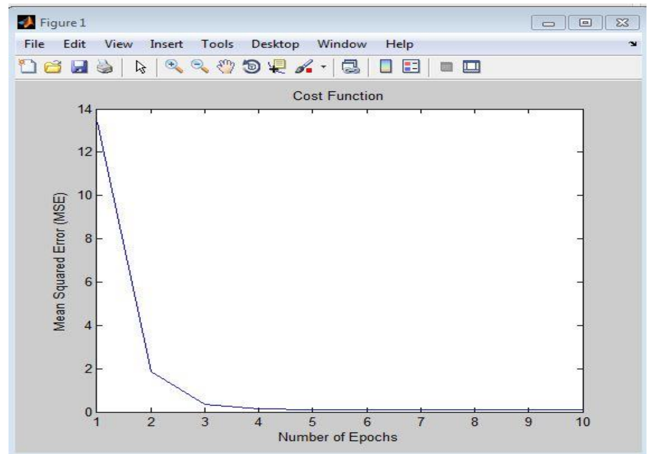


Fig. 6: Number of epochs increase the MSE decreases

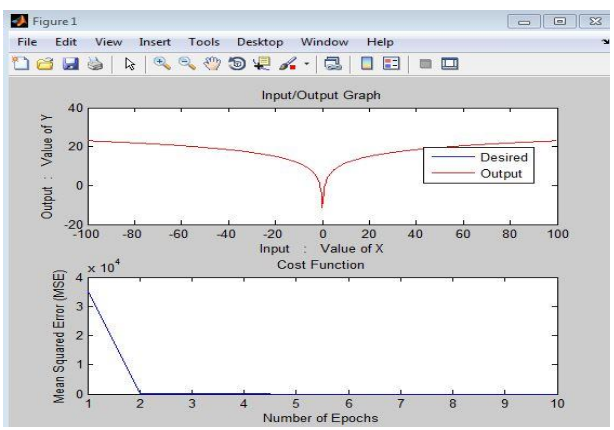


Fig. 7: Number of i/o data value with ref to the mean output value.

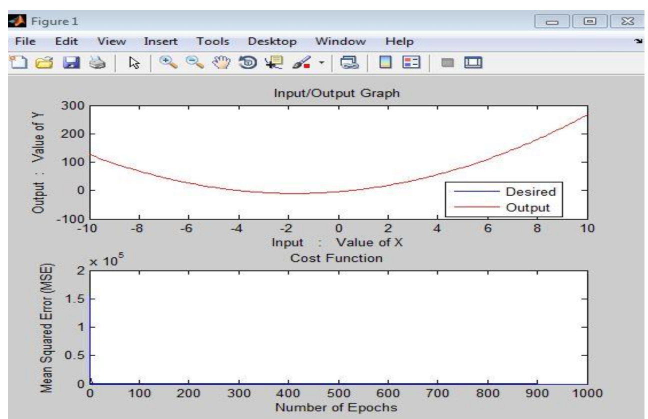
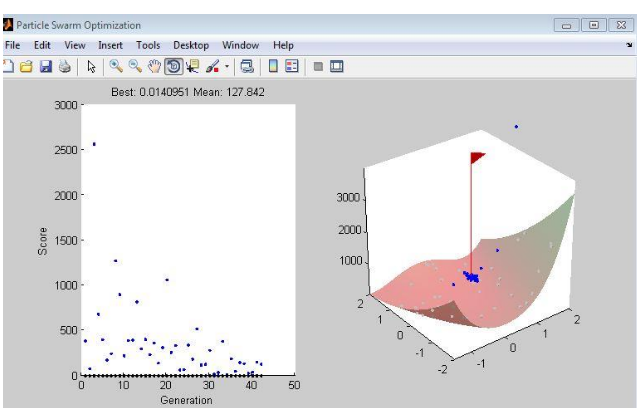


Fig. 8: data driven cost function in data center with number of epochs



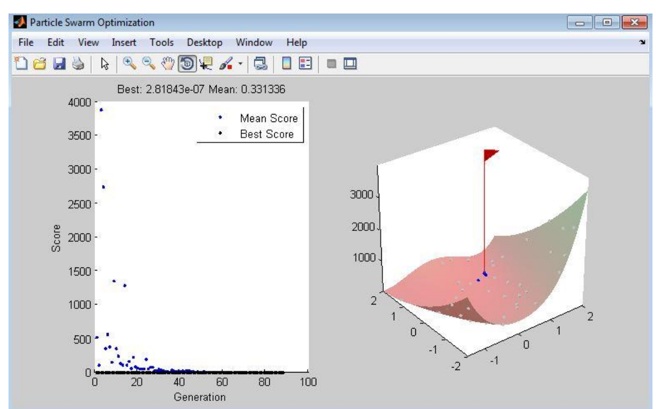


Fig. 9: This is the PSO (right) of scattered data (left)

Fig. 12: This is the PSO-final iteration (right) of scattered data (left)

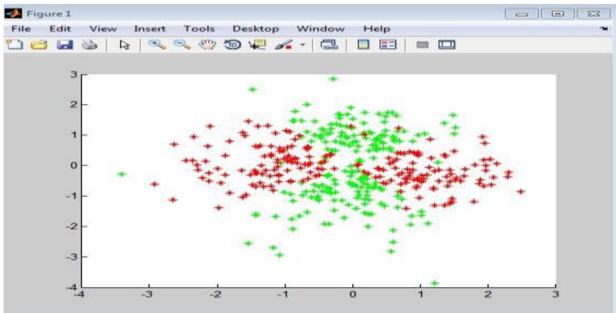


Fig. 10: scattered data set presenting the k-means

Fig. 11 : The above figure shows the execution button for the data processing and encryption

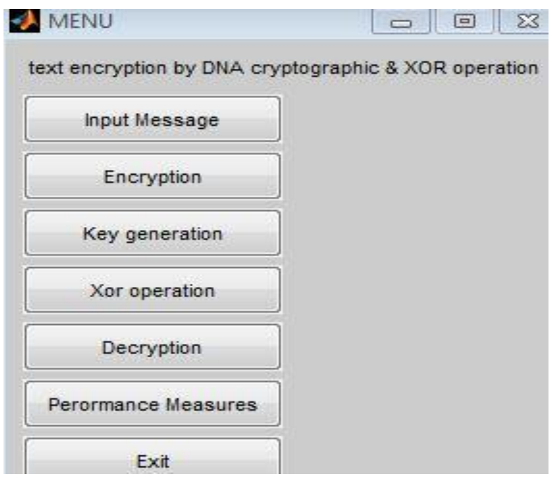


Fig. 12: fig show the process button for cryptography and XOR operation.

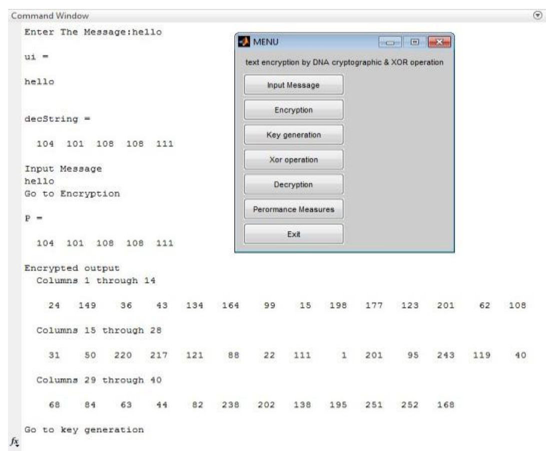


Fig. 13: show the process button for DNA-cryptography process

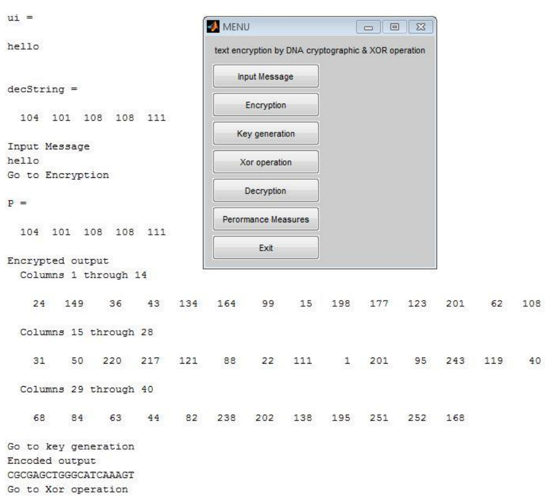


Fig. 14: show the process button for DNA-cryptography generate the key token

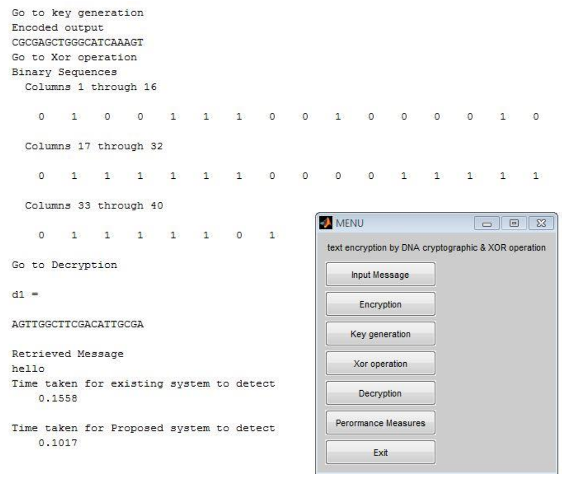


Fig. 15: show the process button for DNA-cryptography show the detection time for hello word.

IV. CONCLUSION

It's easy and straightforward to use. For your organization's institutions, machine learning and artificial intelligence offer tremendous potential. Computational software can make predictions about events or circumstances that occur more quickly than you or your team can deal with, and perhaps work out solutions that are even quicker. Data infrastructure has grown with these new technologies and is a natural extension of current state of data aggregates. Data center efficiency is one of the biggest concerns for IT managers. Increasing performance, reducing, or reducing future problems should be important goals. One should continue to pay attention to the company's assumptions and existing budget. Below, highlighted some of the key issues regarding the efficiency of data centers.

- Barriers to control
- Security
- Cost reduction
- Environmental work

The data center is the economic and strategic organization of the company. However, it is complicated by the multitude of technical areas such as electricity, fire, air conditioning, access control, alarms and often requires many specialized skills. Instead, then needing human involvement or depending on a small set of preset behaviors, machine learning may learn from many situations and data sets and provide instantaneous responses. Managing your data center may be made easier with the use of technology. The shape of the data center is undergoing major changes. To compete in this rapidly evolving digital economy, organizations are deploying workloads in the cloud, clicking on the ecosystem of partners and artificial intelligence leaching (AI) to deliver faster and better digital experiences. These approaches rebuild the underlying infrastructure. No longer in the middle of nowhere, data centers are rapidly disseminated and are ingenious in controlling this extravagant activity. The use case is suitable for machine learning (ML), an AI method that allows the machine to learn and improve itself over time without human input.

References

1. Hazelwood, K., Bird, S., Brooks, D., Chintala, S., Diril, U., Dzhulgakov, D., ... & Wang, X. (2018, February). Applied machine learning at facebook: A datacenter infrastructure perspective. In 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA) (pp. 620-629). IEEE.
2. Chen, L., Lingys, J., Chen, K., & Liu, F. (2018, August). Auto: Scaling deep reinforcement learning for datacenter-scale automatic traffic optimization. In Proceedings of the 2018 conference of the ACM special interest group on data communication (pp. 191-205).
3. Todd, A., Purkayastha, A., Egan, H., Sickinger, D., Eash, M., Serebryakov, S., ... & Phillips, C. (2021). Artificial Intelligence for Data Center Operations (AIOps) (No. NREL/TP-2C00-79712). National Renewable Energy Lab.(NREL), Golden, CO (United States).
4. Berral, J. L., Goiri, Í., Nou, R., Julià, F., Guitart, J., Gavaldà, R., & Torres, J. (2010, April). Towards energy-aware scheduling in data centers using machine learning. In Proceedings of the 1st International Conference on energy-Efficient Computing and Networking (pp. 215-224).
5. Mirhoseini Nejad, S., Badawy, G., & Down, D. G. (2021). Holistic thermal-aware workload management and infrastructure control for heterogeneous data centers using machine learning. Future Generation Computer Systems, 118, 208-218.
6. Haghshenas, K., Pahlevan, A., Zapater, M., Mohammadi, S., & Atienza, D. (2019). Magnetic: Multi-agent machine learning-based approach for energy efficient dynamic consolidation in data centers. IEEE Transactions on Services Computing.
7. Teixeira, P. H. D. S., Clemente, R. G., Kaiser, R. A., & Vieira Jr, D. A. (2010, July). HOLMES: an event-driven solution to monitor data centers through continuous queries and machine learning. In Proceedings of the Fourth ACM International Conference on Distributed Event-Based Systems (pp. 216-221).
8. Tarutani, Y., Hashimoto, K., Hasegawa, G., Nakamura, Y., Tamura, T., Matsuda, K., & Matsuoka, M. (2015, November). Temperature distribution prediction in data centers for decreasing power consumption by machine learning. In 2015 IEEE 7th International Conference on Cloud Computing Technology and Science (Cloud Com) (pp. 635-642). IEEE.