Enhancing Real-Time User IP Tracking and Country Identification with APIs

G V S Abhishek Varma[1], P Srinivasa Rao[2], B Aruna Kumari [3], G V N Akshay Varma[4]

[1,2,3] MVGR College of Engineering, AP, India
[4] SSN College of Engineering, TN, India
abhishekvarma.gvs@gmail.com

Abstract. In the dynamic landscape of modern business, efficiently delivering data across diverse regions with varying demands presents a formidable challenge. Current methodologies, leveraging APIs and machine learning algorithms for real-time user country identification, struggle to adequately optimize resource allocation and data delivery efficiency and calculate it. While IP-based geolocation methods and advanced programming language modules endeavor to track users' IP addresses and identify their countries in real-time, they frequently fail to accurately predict future demand trends and integration with application. Despite efforts to analyze historical data request patterns, existing approaches lack the resilience required to scale resources effectively in specific regions, leading to suboptimal resource utilization and heightened costs for dynamic enterprises such as Over-the-Top (OTT) platforms and payment sites and cloud clients. Prevalent techniques face hurdles in data collection, cleaning, and feature engineering, resulting in inaccuracies and inconsistencies in forecasting future demand trends. As cloud providers track data but won't share to clients in detail which affects clients to build large scale applications commonly employed machine learning algorithms have demonstrate limitations in accurately predicting demand patterns across diverse industries, including e-commerce, healthcare, and finance. The shortcomings of existing technologies in optimizing data delivery efficiency through real-time user IP tracking and country identification highlight the urgent need for an innovative system to effectively address modern data delivery challenges. Therefore, there is a pressing demand for novel approaches that can overcome these limitations and provide more robust solutions for optimizing data delivery and service efficiency in today's dynamic business environment. Such advancements have the potential to significantly enhance resource utilization, reduce costs, and improve the overall user experience, thereby driving innovation and competitiveness across various industries.

Keywords: Geo Tracking, Cloud, Historical Data Analysis.

1 Introduction

The effective delivery of data has become essential for businesses and organizations to thrive. With diverse data requests originating from different regions [1][2], there is

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_18

a pressing need to optimize data delivery in terms of speed, efficiency, and cost-effectiveness. To address this challenge and predict the optimal scaling up of cloud resources in specific regions where requests originate. This endeavor utilizes IP-based geolocation, which enables the identification of the geographical location of internet-connected devices that initiate data requests from a central source. By harnessing the power of IP-based geolocation, determination of the region from which a data request originates, thus facilitating effective resource allocation and finding end clients [3][4][5]. Geolocation accomplishes this by analyzing the unique IP address associated with each device for that individual request, providing valuable insight into the device's location and device type. Leveraging historical data patterns, this way in future we can forecast machine learning algorithms to predict future demand accurately and reliably [6][7]. Machine learning algorithms are particularly advantageous in this context as they can process extensive datasets, uncover hidden patterns, and derive valuable insights and may include traditional methods. The ultimate objective is to enhance data delivery performance and efficiency [8][9]. By accurately predicting future demand, resources can be allocated optimally, enabling the prompt and cost-effective delivery of data while ensuring its quality remains uncompromised. One exemplary use case lies within the realm of cloud-based storage providers operating across various global regions. Typically, these providers store data in a central location and distribute it to users based on their geographical location. However, as demand for data fluctuates dynamically across regions, the provider may need to adjust the scaling of their VMs to accommodate changing demands effectively. By leveraging machine learning algorithms to predict future demand, cloud-based storage providers can optimize resource allocation, guaranguarantee swift and efficient data delivery [10][11]. For instance, if it predicts a substantial increase in data requests from a particular region, the provider can proactively scale up their VMs in that region, ensuring prompt and uninterrupted data delivery and doesn't depend on cloud provider. This intelligent approach pro-motes efficient utilization of resources, preventing unnecessary expenditures [12][13]. The applicability of this method extends beyond cloud-based storage providers, finding utility in various industries and use cases [14][15]. Online retailers can leverage this approach to predict product demand in different regions, enabling optimal resource allocation. Similarly, news websites can anticipate the demand for news articles across diverse geographic areas and add languages to that article, fine-tuning their servers to deliver content with exceptional speed and efficiency. By revolutionizing the way data is delivered and managed across different regions and industries, the approach holds significant potential for the advancement of prediction and artificial intelligence applications [16][17][18].

2 Related Work

The use of machine learning (ML) algorithms for predicting future demand and optimizing resource allocation is a popular research topic in the field of cloud computing. Several studies have explored the use of ML algorithms for cloud resource management, including VM scaling, load balancing, and resource provisioning and even in the field of data science they are ways that use this to track users location and improve the user experience and prices in that locality based on the count of views[19][20][21] and the recent trend of calculating royalty for movies. Study by Thangavel et al. (2020) proposed a resource management framework for cloud-based applications using ML algorithms [22]. The study used historical data to train an ml model that predicted future demand and optimized resource allocation. The results showed that the proposed framework improved resource utilization and reduced operational costs [23]. Another study by Mirakhorli et al. (2018) explored the use of ML algorithms for load balancing in cloud computing. The study proposed a hybrid algorithm that combined clustering and classification techniques to predict future workload and allocate re-sources accordingly [24]. The results showed that the algorithm outperformed traditional load balancing algorithms in terms of resource utilization and response time. The use of geolocation for optimizing resource allocation is also a popular research topic. Several studies have explored the use of geolocation data to identify the location of users and optimize resource allocation accordingly. One study by Wang et al. (2018) proposed a geolocation-based VM scaling approach for cloud-based applications [25][26]. The study used geolocation data to identify the location of users and predict future demand. The proposed approach improved resource utilization and reduced operational costs [27][28]. Another study by Chen et al. (2019) explored the use of geolocation data for dynamic VM provisioning in cloud computing [29][30]. The results showed that the proposed framework improved resource utilization and reduced operational costs and the increase of applications based on location and restrictions this type of Application can be helpful [31][32]. The application of ML algorithms for optimizing re-source allocation in cloud computing has several benefits, including reducing operational costs, improving resource utilization, and enhancing the overall performance of the system. However, there are also some challenges associated with the use of ML algorithms, such as the need for large amounts of historical data, selecting the appropriate algorithm, and ensuring the reliability and accuracy of the predictions. To address these challenges, several studies have proposed novel approaches, such as using hybrid algorithms, combining different techniques, and improving the accuracy of geolocation data. For example, a recent study by Kumar et al. (2021) proposed a hybrid algorithm that combined deep learning and reinforcement learning for resource allocation in cloud computing. The study used historical data to train the algorithm and demonstrated that the proposed approach outperformed traditional methods in terms of resource utilization and response time [33]. The article predicts the optimal scaling up of a virtual machine (VM) in a particular region based on the previous data requests from different regions is a valuable contribution to the field of cloud computing. By using IP-based geolocation and machine learning, the project can optimize resource allocation, improve the performance of the system, and reduce operational costs. The example use case for a cloud-based storage provider is just one of many potential applications for this approach [34]. As research in this field continues to evolve, we can see advancements of ML algorithms and geolocation data for cloud resource management.

3 Methodology

It begins with user interactions on the client's website/application, triggering an API call to Azure services to retrieve their IP addresses with time stamp. Azure's robust infrastructure processes these requests efficiently. Subsequently, retrieved IP addresses are stored in Azure Storage within a CSV file. If the CSV file is non-existent, it is dynamically created with an IP address header. The data is then accessed and analyzed in an Azure ML Notebook, where comprehensive data operations take place. This includes data cleaning and preprocessing to ensure high data quality and consistency. Statistical analysis techniques are employed to discern user behavior patterns, session durations, and other pertinent metrics. Additionally, geolocation data is leveraged to map IP addresses, providing in-sights into geographical distribution patterns. Finally, data visualization techniques are implemented to effectively communicate research findings, enabling stakeholders to make informed decisions. This methodology ensures a systematic and data-driven approach to understanding user behavior and enhancing the applications performance.

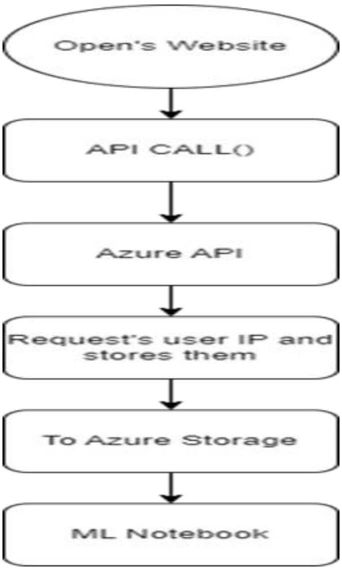

Fig. 1. API'S Working

Algorithm:

Step 1: User Accesses Website:

- When a user accesses the website, the algorithm initiates.

Step 2: API Call to Azure Services:

- The website sends an API call to Azure services to retrieve the user's IP address and API process and returns IP address.

Step 3: Store IP Address in Azure Storage:

- The algorithm stores the user's IP address in Azure Storage. If the CSV file doesn't exist, it creates a new one and adds the IP address as a new record.

Step 4: Azure ML Notebook:

- Load the CSV file with user IP addresses from Azure Storage to Azure Notebook. Perform data cleaning and preprocessing.
- Conduct statistical analysis to identify patterns, trends, and geolocation information.

4 Environmental Setup

The proposed methodology employs a synergistic blend of tools and technologies to establish an efficient data delivery system centered around IP-based geolocation and machine learning capabilities. This robust setup is instrumental in extracting valuable insights from user interactions. Azure Function App serves as the foundational building block, providing a serverless computing service that simplifies application development and deployment. This approach allows developers to allocate their focus towards application logic rather than grappling with infrastructure management. It enables rapid responsiveness to various events, including HTTP requests like Get and post and messages from Azure Event Grid or Azure Service Bus. To orchestrate data processing and IP-based geolocation, Python, a versatile and widely-used programming language, is harnessed in conjunction with Azure Function App. Python's flexibility empowers developers to seamlessly integrate essential modules such as logging, azure.functions, csv handling, and datetime management into the Azure Function. Azure Machine Learning Studio plays a pivotal role in handling extensive datasets and facilitating machine learning tasks. This cloud-based service empowers developers to create, deploy, and manage machine learning models with ease. Its scalability, adaptable computing resources, and support for various file formats, including Py, ipynb, and csv, enhance the project's analytical capabilities and directly synchronize with the main live dataset. In the realm of IP-based geo-location, MaxMind GeoIP2 Database (MMDB), a publicly accessible repository developed by MaxMind. This database contains comprehensive data for both IPv4 and IPv6 addresses. Leveraging the geoip2 module in Python, the project seamlessly interfaces with the MMDB, enabling precise IP-based geolocation queries. This database serves as an invaluable resource for developers seeking to incorporate geolocation functionalities into their applications. Collectively, this environmental setup forms a robust foundation for efficient data delivery, advanced geolocation analysis, and machine learning insights, all while ensuring a seamless and responsive user experience. These tools and technologies are combined to implement the proposed method effectively. Azure Function App provides the event-driven infrastructure, Python handles data processing and geolocation, Azure Machine Learning Studio supports machine learning tasks, and the MMDB database enables accurate IP-based geolocation. The integration of these components enables efficient resource allocation and data delivery. By leveraging these tools and technologies, the method optimizes data delivery by accurately identifying the geographical location of users through IP-based geolocation. The combination of Azure Function App, Python, Azure Machine Learning Studio, and the MMDB database forms a powerful framework for achieving efficient data delivery while minimizing infrastructure management concerns. User's Ip address and there visiting timestamp in GMT.

Results and Discussions

Used azure functions apps, and a specific function was created within the app, opting for the "HTTP trigger" template. To handle incoming HTTP requests and store the relevant data, a Python script was implemented. The script enabled the saving of data in a CSV file format. Once the function, named "HTTPtrigger1," was deployed to the Azure Functions app, it underwent rigorous testing using various test inputs, including "get" and "post" requests. This comprehensive approach facilitated the generation of a robust dataset by utilizing an HTTP trigger mechanism and leveraging the capabilities of Python programming and then calculations based on it showing the top languages user played based on their location and best place to host a CDN and storage options.

	A	B
1	Country	Visitors
2	India	49.204.229.143:39691,2024-01-31 05:44:58
3	India	49.204.229.143:39767,2024-01-31 05:45:22
4	India	49.204.229.143:39767,2024-01-31 05:45:27
5	India	49.204.229.143:39767,2024-01-31 05:45:30
6	India	49.204.229.143:39936,2024-01-31 05:47:58
7	India	49.204.229.143:40111,2024-01-31 05:50:40
8	India	49.204.229.143:40111,2024-01-31 05:50:43
9	India	49.204.229.143:40111,2024-01-31 05:50:43
10	India	49.204.229.143:40111,2024-01-31 05:50:43
11	India	49.204.229.143:40111,2024-01-31 05:50:44
12	India	49.204.229.143:40111,2024-01-31 05:50:44
13	India	49.204.229.143:40254,2024-01-31 05:53:42

Fig. 2. Live Dataset

The data analysis process to ensure that the data is accurate and reliable. used panda's module to read the CSV file containing IP address data and extract the first column containing the country names. We then use the mmdb module to obtain the country details for each IP address and save it to a CSV file. To ensure that the data is organized and structured properly, we give column names to each column. The first column is named "Country" and contains the country names, while the remaining columns are named after the IP addresses. We can easily ex-tract the user count for each country by looping through the remaining columns and updating the count of users for each country. The use of appropriate column names and modules like pandas and mmdb can help to ensure that the data is accurate and reliable and calculating country wise users and predict their language using seaborn (See fig.3, fig.4)

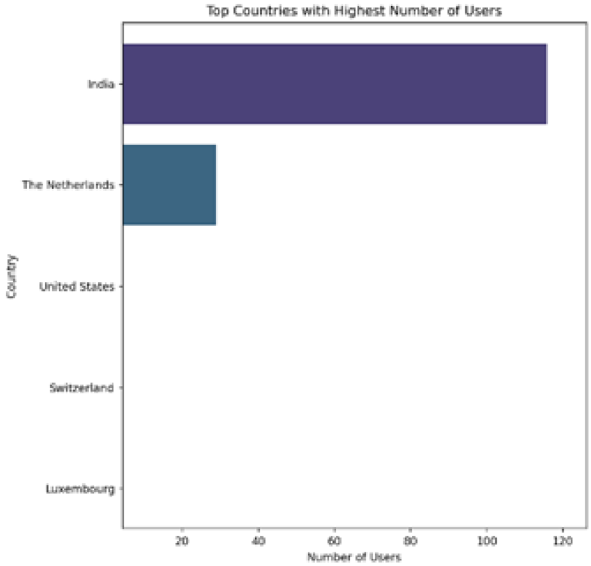

Fig. 3. Top Country Wise Users

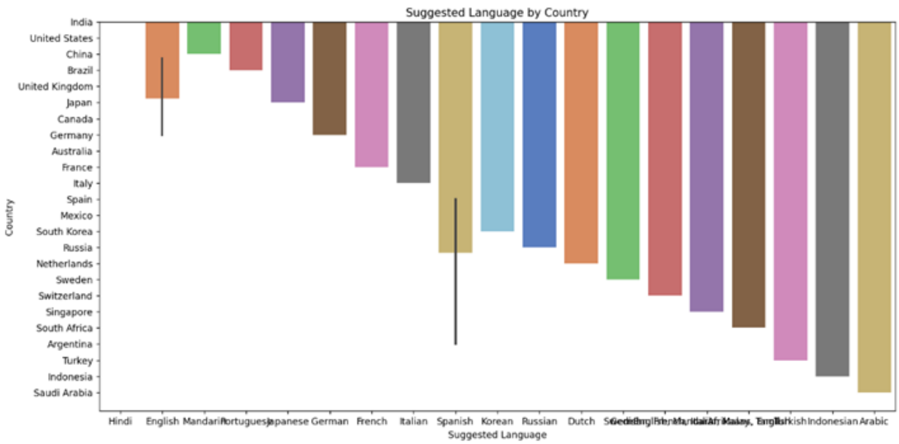

Fig. 4. Language by Country

5 Conclusion

To predict the optimal scaling up of a cloud resource using IP-based geolocation and machine learning algorithms has the potential to improve the performance and efficiency of data delivery in cloud computing. By analyzing historical patterns of data requests, the project can predict future demand and allocate resources efficiently to ensure that data is delivered quickly and at a reasonable cost. The use of geolocation data to identify the location of users is a powerful tool for optimizing resource allocation and ensuring that data is delivered to users in a timely manner. The project's use of Azure Functions to record users' IP addresses and train machine learning models for predicting future demand is a practical and effective approach to optimizing resource allocation in cloud computing. By leveraging Azure Functions, the project can easily scale up or down its resources as needed to handle varying demand. Overall, the project has the potential to benefit cloud-based storage providers and other organizations that need to deliver data quickly and efficiently to users in different regions and keep it secure. By improving resource allocation and ensuring that data is delivered in a timely manner. The project can help organizations to provide better service to their customers while reducing operational costs and can be helping security experts to log data and for applications that can track user IP address to allow them to access the application and predict them. It helps OTT platforms to build this type of model to predict users and save costs.

References

1. Fruchter N, Miao H, Stevenson S, Balebako R. Variations in tracking in relation to geographic location. arXiv preprint arXiv:1506.04103. 2015 Jun 12.
2. Xie, Yinglian, et al. "How dynamic are IP addresses?." Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications. 2007.
3. Forell, Tim, Dejan Milojicic, and Vanish Talwar. "Cloud management: Challenges and opportunities." 2011 IEEE international symposium on parallel and distributed processing workshops and Phd forum. IEEE, 2011.
4. S Satyanarayana ,"Privacy Preserving Data Publishing Based On Sensitivity in Context of Big Data Using Hive",Journal of Bigdata(Springer), Volume:5,Issue:20, ISSN: 2196-1115, July 2018.
5. Sreenivas, V., B. ArunaKumari, and J. VenkataRao. "Enhancing the security for information with virtual data centers in cloud." Future Wireless Networks and Information Systems: Volume 1. Springer Berlin Heidelberg, 2012.
6. P.Mahesh Kumar,P. Srinivasa Rao, "Frequent Pattern Retrieval on Data Streams by using Sliding Window", EAI Endorsed Transactions on Energy web,Volume:5,issue:35,2021.
7. Sim, Kwang Mong. "Agent-based cloud computing." IEEE transactions on services computing 5.4 (2011): 564-577.
8. T.V. Madhusudhana Rao, Suresh Kurumalla, Bethapudi Prakash, "Matrix Factorization Based Recommendation System using Hybrid Optimization Technique, EAI Endorsed Transactions on Energy Web,Volume:5,issue:35,2021.

9. Thangavel, K., & Manickam, S. (2020). Resource management in cloud computing using machine learning: A review. Journal of Ambient Intelligence and Humanized Computing, 11(2), 807-824. doi: 10.1007/s12652-019-01467-3
10. Mirakhorli, M., Zamanifar, K., & Wong, K. (2018). Machine learning-based load balancing in cloud computing: A survey. IEEE Access, 6, 30476-30494. doi: 10.1109/ACCESS.2018.2836941P3
11. S.Vidya sagar Appaji, P. V. Lakshmi, "Maximizing Joint Probability in Visual Question Answering Models", International Journal of Advanced Science and Technology Vol. 29, No. 3, pp. 3914 – 3923,2020.
12. Vidya sagar Appaji setti ,P Srinivasa Rao , "A Novel Scheme For Red Eye Removal With Image Matching", Journal of Advanced Research in Dynamical & Control Systems, Vol. 10, 13-Special Issue, 2018.
13. Krishna Prasad, P.E.S.N, "A Secure and Efficient Temporal Features Based Framework for Cloud Using MapReduce", springer, 17th International Conference on Intelligent Systems Design and Applications (ISDA 2017),Volume:736,pp:114-123, ISSN 2194-5357 Held in Delhi,India, December 14–16, 2017.
14. Wang, Y., Ye, X., Wang, J., & Huang, J. (2018). Geolocation-based virtual machine scaling for cloud applications. Journal of Parallel and Distributed Computing, 116, 20-30. doi: 10.1016/j.jpdc.2018.01.004
15. Chen, T., Li, Z., Li, Q., Wang, Y., & Zhang, Y. (2019). A dynamic virtual machine provisioning framework based on user geolocation in cloud computing. Journal of Parallel and Distributed Computing, 123, 162-176. doi: 10.1016/j.jpdc.2018.10.011
16. Madhusudhana Rao, T.V., Srinivas, Y, "A Secure Framework For Cloud Using Map Reduce",Journal Of Advanced Research In Dynamical And Control Systems(IJARDCS), Volume:9,Sp-14,Pp:1850-1861,ISSN:1943-023x,Dec, 2017.
17. Bharathi Uppalapati, S Srinivasa Rao," Application of ANN Combined with Machine Learning for Early Recognition of Parkinson's Disease", Intelligent System Design: Proceedings of INDIA 2022, Springer Nature Singapore,PP: 39-49, 28/10/2022.
18. Sushma Rani N, "An Efficient Statistical Computation Technique for Health Care Big Data using R", Scopus, IOP Conference Series: Materials Science and Engineering, Volume: 225, ISSN:1757-8981,ISSUE NO :012159,2017.
19. J. Zhu, P. He, Z. Zheng and M. R. Lyu, "A Privacy-Preserving QoS Prediction Framework for Web Service Recommendation," 2015 IEEE International Conference on Web Services, New York, NY, USA, 2015, pp. 241-248, doi: 10.1109/ICWS.2015.41.
20. G V S, Abhishek & Kethaan, Adari & Akshay Varma, G V N. (2023). Predictive Insights for Monthly Property Sales Forecasting: An End-to-End Time Series Forecasting. International Journal of Advanced Research in Science Communication and Technology. 3. 10.48175/IJARSCT-12493.
21. Abhishek Varma G V S, Akshay Varma G V N."Dynamic User Routing for Paid and Free Users in Web Applications using Content Delivery Network (CDN)", Volume 11, Issue VII, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 409-414, ISSN : 2321-9653, www.ijraset.com
22. Shukla, P. K., Maheshwary, P., Subramanian, E. K., Shilpa, V. J., & Varma, P. R. K. (2023). Traffic flow monitoring in software-defined network using modified recursive learning. Physical Communication, 57, 101997.
23. Kingo, Thomas, and Diego F. Aranha. "User-centric security analysis of MitID: the Danish passwordless digital identity solution." Computers & Security 132 (2023): 103376.

24. Biryukov, Alex, and Sergei Tikhomirov. "Deanonymization and linkability of cryptocurrency transactions based on network analysis." 2019 IEEE European symposium on security and privacy (EuroS&P). IEEE, 2019.
25. Priya, VS Devi, and S. Sibi Chakkaravarthy. "Containerized cloud-based honeypot deception for tracking attackers." Scientific Reports 13.1 (2023): 1437.
26. G V S, Abhishek & Kethaan, Adari & Akshay Varma, G V N. (2023). Fraud Detection System Employing Machine Learning Techniques for Credit Card Transactions. International Journal of Advanced Research in Science Communication and Technology. 3. 2581-9429. 10.48175/IJARSCT-12492.
27. Böck, Leon, et al. "How to Count Bots in Longitudinal Datasets of IP Addresses." NDSS. 2023.
28. S. Arvind, S. Arvind, V. K. Silveri, G. Potey, P. Nunavath and R. Podishetty, "Network Traffic Virtualization Using Wireshark and Google Maps," 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballar, India, 2023, pp. 1-6, doi: 10.1109/ICDCECE57866.2023.10150823.
29. Krishna Prasad, M.H.M., Thammi Reddy, K , “A Efficient Data Integration Framework in Hadoop Using MapReduce" Published in Computational Intelligence Techniques for Comparative Genomics , Springer Briefs in Applied Sciences and Technology , ISSN:2191-530X,PP 129-137, October 2014
30. Nagesh Vadaparhi, Srinivas Yarramalle, “A Novel clustering approach using Hadoop Distributed Environment”, Springer, (Applied Science and Technology), ISSN:2191-530X,Volume:9, pp:113-119,October 2014
31. Balajee Maram, Guru Kesava Dasu Gopisetty, “A Framework for Data Security using Cryptography and Image Steganography”,International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-11,September,2019.
32. Chen, Pengzhan, Jihua Wu, and Ning Li. "A personalized navigation route recommendation strategy based on differential perceptron tracking user’s driving preference." Computational intelligence and neuroscience 2023 (2023).
33. Yen, Ting-Fang, et al. "Host Fingerprinting and Tracking on the Web: Privacy and Security Implications." NDSS. Vol. 62. 2012.
34. Almohri, Hussain MJ, Layne T. Watson, and David Evans. "Predictability of IP address allocations for cloud computing platforms." IEEE Transactions on Information Forensics and Security 15 (2019): 500-511.