International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Geospatial Analytics in AWS Cloud Using AWS SageMaker

Venkat Soma

New York Mets

Abstract: ***This paper will notably explore the implementation of geospatial analytics in AWS utilizing Amazon Sagemaker, showing its transformative effects on different industries. PageMaker's capabilities effectively enable the effective development, deployment, and training of machine learning models along with geospatial data through addressing conventional data management challenges. The key characteristics involve earth observation jobs, vector enrichment jobs, and map visualization that increase decision - making and insights. It had several numbers for applications involving climate monitoring, agriculture, disaster management beside urban development.***

Keywords: Amazon SageMaker, Amazon Sagemaker Studio Classic, earth observation jobs, vector enrichment jobs, and map visualization

1. Introduction

a) Project Specification

Geospatial analytics in AWS cloud utilizing the Amazon SageMaker is a crucial system that fosters geospatial data such as location data, maps, and satellite imagery for developing, training, and deploying machine learning (ML) models. In short, Amazon SageMaker's geospatial capabilities effectively allow individuals to create, terrain and deploy ML models utilizing geospatial data [1]. The data can notably enrich or transform large - scale geospatial data sets. A person can encourage model building with a pre - trained machine - learning model and recognize model prediction within an interactive map utilizing 3D accelerated graphics besides a built - in visualization tool.

b) Aim and Objectives

Aims: This project aims to attach Amazon Sagemaker's geospatial data for diversified applications in disaster response and food security.

Objectives:

- To develop and train the machine learning models using geospatial data for various applications.
- To create and visualize the earth observation strategies for evaluating and mitigating the effect of natural disasters.
- To enhance the strategies for urban development by analyzing the geospatial information.

c) Research Question

RQ1: How can Amazon SageMaker's geospatial analytics be utilized to improve disaster response, food security, and urban development?

RQ2: What is the significance of geospatial analysis and ML models to optimize resource management and decision - making processes?

d) Research Rationale

The geospatial data could be utilized for different kinds of use cases involving natural disaster responses and management, improving food security, and enhancing the harvesting yield. It further helps to support efficient and sustainable urban development by providing several benefits. Through visualizing and creating an earth observation strategy for the land cover segmentation, an organization could assess the loss of harvest mainly caused by wildfires. The organization can significantly mitigate the damage by effective observation and analysis.

2. Literature Review

a) Research background

In this recent time, the overall requirement for timely and precise geospatial data analysis has been suddenly increasing throughout different industries such as agriculture services, financial services, and financial and real estate services. Conventional methods regarding the management of geospatial data are sometimes unmanageable for several kinds of reasons [2]. This mainly happens due to long model training periods, large - scale data processing, and complex infrastructure setups. These specific challenges hamper the ability to make critical decisions. It also hinders to achieve actionable insights. However, the incorporation of diverse data sources such as open - source datasets and existing consumer data poses some additional technical challenges.

b) Critical Assessment

The absence of specialized data tools for effective geospatial data management and the requirement for custom visualization worsen the above - mentioned issues. Hence, as an outcome business organizations face different kinds of issues such as increasing costs, and inefficiencies, besides a much slower time to market for their geospatial data analysis and making critical decisions. Aside from this, agriculture also faced issues such as ineffective yield prediction, disease management, and poor crop monitoring due to limited access to accurate and timely geospatial data [3]. Similarly, conventional methods are more of a slow and labor - intensive which significantly leads to the suboptimal decision.

c) Linkage to Aim

This system's capabilities primarily aim to address the significant problem by providing effective access to geospatial data and purpose - built algorithms for data model training and data preparation. These capabilities effectively decrease the technical barriers to interpreting and generating geospatial prediction. Organizations can utilize SageMaker capabilities for making predictions on the geospatial data faster compared to the traditional solutions. SageMaker's

geospatial abilities make it much easier to access geospatial data through the existing consumer datasets, other SageMaker geospatial data providers, and open - source data sets.

d) Encapsulation of applications

In this concern, the Amazon SageMaker geospatial capabilities could solve the above - mentioned issues by providing significant access to pre - trained models and geospatial data. Along with a purpose - built algorithm for data analysis and preparation, farmers can effectively monitor crop health, handle diseases and pests, and predict yield [4]. The overall capability of this system would decrease the requirement for developing data preprocessing functions and customer infrastructure by offering purpose - built algorithms. This provided algorithm by Amazon would effectively help for effective data preparation, interference, and model training.

e) Utilizing geospatial for visualization

There are three key "Amazon SageMaker geospatial capabilities" which involve earth observation jobs, vector enrichment jobs, and map visualization. In this earth observation job, an individual can transform, obtain, and pre - train ML models for developing a prediction and achieving valuable insights. The "vector enrichment jobs" significantly enrich the data with operations such as transforming geographical coordinates towards a readable address [5]. The map visualization helps to properly visualize the map data and satellite images uploaded through a GeoJSON, JSON, or CSV file [6]. An individual can also develop an earth observation job within the SageMaker studio notebook for effectively processing the satellite data utilizing a purpose - built geospatial operation. Some of the list of "purpose - built geospatial operations" mainly supported by the SageMAker studio involve cloud masking, cloud renewal, and band stacking.

f) Theoretical Framework

Studio Classic UI is a studio classic notebook that utilizes geospatial 1.0 images [7]. organizations can implement a purpose - built SageMaker geospatial image that supports both GPU and CPU - based notebook instances. They can significantly execute the initially utilized open - source libraries which are found in the geospatial machine learning streamlined works. Individuals can use the Amazon Sagemaker processing beside the geospatial container for implementing a large - scale workload with the owned datasets involving satellite imagery, commercial application, LiDAR, climate, weather, and soil. The utilization of earth observation jobs can help an individual optimize data processing.

g) Literature Gap

The literature sheds light on the challenges of traditional geospatial data analysis but lacks comprehensive insights into the specific benefits and applications of Amazon SageMaker's geospatial capabilities.

3. Methodology

a) Research Philosophy

This project is based on the overview of the utilization of the Geospatial Analytics in AWS Cloud Using AWS SageMaker. The research will use the philosophy of interpretivism to emphasize the perspectives of researchers on this topic. It will explore the opinions of the developers, users, and administrators about the significance of geospatial analysis within a cloud system. Interpretivism encompasses social theories and perspectives that embrace a view of reality as socially constructed.

b) Research Approach

The research regarding the complete analytical framework of geospatial systems in disaster management through the development of software tools and techniques in agricultural sectors. This research includes the deductive approach to investigate the efficacy of the integration of Amazon Sagemaker with continuous deployment for automated testing of business organizations. By using the deductive method this project will provide the opinion of previously working individuals through data collection and analysis.

c) Research design

To collect and analyze the data about the performance of Sagemaker within the actions of the agricultural and commercial purposes, the secondary qualitative method is used. It will help to provide an overview of the development and deployment process within the Geospatial analytics in the AWS cloud.

d) Data collection method

The data collection will be practiced through peer review of previously published scholarly articles, and journals accessed through Google Scholar and PubMed. The obtained information will be accumulated and analyzed based on thematic analysis.

e) Ethical considerations

In this study, the maintenance of the ethical perspectives is one of the most significant sections. Firstly, privacy and permission laws must be followed when using confidential information about sports events.

4. Results

a) Critical Analysis

There are three key "Amazon SageMaker geospatial capabilities" which involve earth observation jobs, vector enrichment jobs, and map visualization. In this earth observation job, an individual can transform, obtain, and pre - train ML models for developing a prediction and achieving valuable insights. The "vector enrichment jobs" significantly enrich the data with operations such as transforming geographical coordinates towards a readable address. The map visualization helps to properly visualize the map data and satellite images uploaded through a GeoJSON, JSON, or CSV file [8]. An individual can also develop an earth observation job within the SageMaker studio notebook for effectively processing the satellite data utilizing a purpose - built geospatial operation. Some of the list of "purpose - built geospatial operations" mainly supported by the SageMAker studio involve cloud masking, cloud renewal, and band stacking.

b) Finding and Discussion

Theme 1: Enhanced Agricultural Management through Geospatial Data

Business organizations and farmers can effectively monitor yield prediction and crop health besides managing the disease and pests. Individuals can develop and share custom data and visualization with the company from the Amazon Sagemaker Studio Classic. The geospatial capabilities of SageMaker would provide a pre - trained model regarding the common utilization in financial service, insurance, real estate, and agriculture.

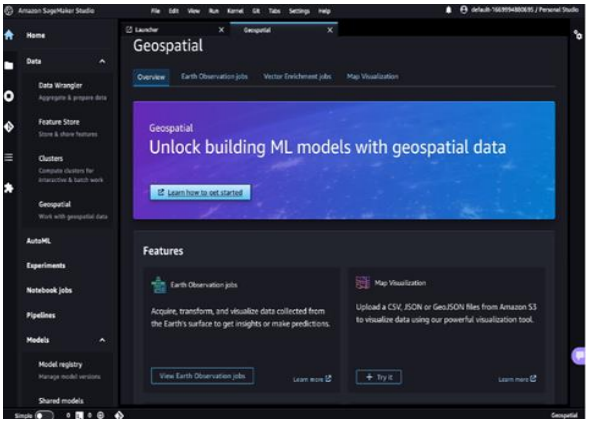

Figure 1: Geospatial machine learning features [13]

Digital farming consists of implementing digital solutions to help farmers optimise crop production within the agriculture sector by the utilization of advanced machine learning and analytics [9]. In this concern, application of the digital farming significantly requires properly working with the geospatial data involving s satellite imagery field in which farmers harvest their overall foods. A farmer can implement Sagemaker for identifying the farm field boundaries within the satellite imagery by a pre - trained model regarding the land cover classification [10]. Individuals can enhance productivity by utilizing pre - trained models and purpose - built algorithms for speeding up the model training and building. An organization or person can significantly utilize the built - in visualization tools for recognizing the prediction output on a communicative map and then collaborate throughout the teams [11]. Utilizing the visualization characteristics provided by the Amazon SageMaker geospatial, organization can properly visualize the data. In the following image, the capabilities of the visualization tool powered by SageMaker geospatial have been depicted.

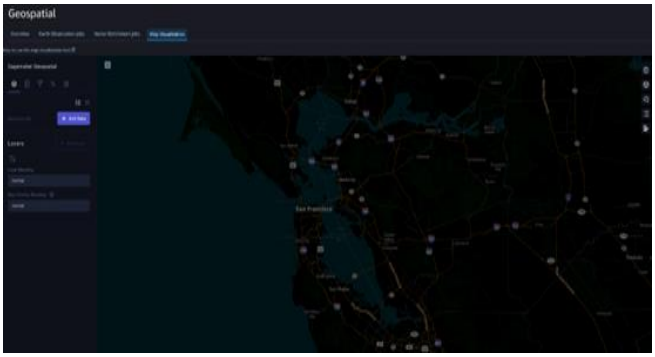

Figure 2: Visualization capabilities of the SageMaker geospatial [7]

Consumers throughout different industries are utilizing the amazon sagemaker geospatial capabilities for real - world implementation. This system provides several impacts which are given in the below section

Theme 2: Disaster Response and Damage Assessment

In recent times, natural diseases have significantly increased. As the severity and frequency of the natural disaster is significantly increasing hence it is vital to equip first responders and decision - makers with accurate and faster damage assessment [12]. An individual can utilize geospatial imagery to predict natural diseases among information and vital data regarding the aftermath of natural data. This data can help to significantly recognize the damage to roads, buildings and other vital infrastructure. Through an example notebook, an individual can train, predict and deploy natural disaster damages through the floods of Rochester, Australia during mid - October 2022 [13]. In the below figure, a specific image from before and after the disaster was taken being the output of the trained machine - learning model. The outcome of the segmentation mask regarding the Rochester floods has been depicted below. In these images, it can be observed that the model had effectively identified the location within the flooded area that would be likely to get damaged. Organizations can both deploy and train a geospatial segmentation model for assessing wildfire damages utilizing the "multi - temporal sentinel - 2 - 2 satellite data" by the GitHub repository. The above area location is an example that is located in Northern California.

Figure 3: Pre - flood and post - flood [13]

Theme 3: Urban Development and Climate Monitoring

Recently, climate change has been a serious concern for the environment. Earth's overall climate change enhances the risk and challenges of droughts the reason for global warming [14]. Hence, individuals or organizations can actively data, implement analysis and effectively visualize the change along with the Sagemaker geospatial abilities for scrutinizing the shrinking shoreline due to climate change. The advanced notebook example elaborates on the procedure of utilizing the Sagemaker geospatial abilities for performing a vector - based map - matching operation and effectively visualizing the outcome. The process of map matching would allow for noisy and snap GPS coordinates toward road segments. Along with the Amazon sagemakers' geospatial abilities, it might be possible or perform a VEJ regarding the map marching. These kinds of jobs grasp a CSV file along with the route information involving GPS measurement timestamps, latitude, and longitude as crucial input and generate a

GeoJSON file which consists of the predicted route. Utilization of digital technologies such as Machine learning helps to explore the effects of heat among urban areas and the different aspects that influence the local temperature for delivering a better design besides supporting sustainable results [15]. However, the urban heat island besides its associated discomforts and risks is one of the biggest challenges that the metropolitan cities are facing today. In this recent time, utilizing Amazon's Sagemaker geospatial capabilities helps to find out the urban heat elements by the earth observation data that significantly provided several benefits and advantages.

c) Evaluation

Individuals or organizations can foster Amazon Sagemaker's pre - trained models beside purpose - built algorithms for improving disease management, yield prediction and crop monitoring in agriculture. Moreover, it would help in implementing natural disaster damage, increasing urban development planning, and monitoring climate change. The system had a significant ability to incorporate diverse data sources and give a custom visualization which effectively increase its utility, making it an indispensable strategy for geospatial data analysis. Business organizations might foster improved geospatial analytics for optimizing the supply chain predicting consumer demands and effectively planning the retail location.

5. Conclusion

Customers throughout different industries are utilizing the amazon sagemaker geospatial capabilities for real - world implementation. There are several benefits this system provides which involve enhancing food security and harvest yield, damage assessment, monitoring climate change, predicting the demands of retail, and supporting and improving urban development. The advanced processing tools and data analysis would lead to precise climate models besides real - time analytics for natural disaster management. In conclusion, it can be stated that Amazon SageMaker's geospatial capabilities effectively transform geospatial data management and utilization providing powerful solutions and tools for addressing the contemporary issues throughout the different sectors.

6. Research Recommendations

The incorporation of geospatial analytics in AWS, particularly by the Amazon Sagemaker has revolutionized the utilization and handling of geospatial data. Hence, by enabling the deployment, training and development of this machine learning model utilizing geospatial data, sagemaker mitigates several challenges encountered by the conventional geospatial data management strategy. It effectively streamlined the overall procedures, increased efficiency and reduced costs throughout the different industries involving urban development, finance and agriculture. SageMaker's overall capabilities, such as earth observation jobs, vector enrichment jobs, and map visualization give a comprehensive method for visualizing and transforming geospatial data, thus facilitating critical decision - making and actionable insights.

7. Future Work

The future scope of the Amazon Sagemaker's geospatial capabilities is effectively expanding and promising a notable advancement throughout the different sectors by the incorporation of geospatial data along with machine learning. Some of the potential future scope is given in the below section. Future development might enable more accurate management and monitoring systems regarding irrigation requirements, soil conditions and crop health [17]. This advanced model could effectively predict the harvesting time and optimal planning for reducing resource usage and improving yield.

References

[1] Amazon Web Services, Inc. . Geospatial Data Science - Amazon SageMaker Supports Geospatial ML Features - Amazon Web Services. [online] Available at: https: //aws. amazon. com/sagemaker/geospatial/features/ [Accessed 24 Jul.2024].

[2] Durante, K. and Hardy, D., 2015. Discovery, management, and preservation of geospatial data using hydra. *Journal of map & geography libraries*, *11* (2), pp.123 - 154. https: //www.tandfonline. com/doi/abs/10.1080/15420353.2015.1041630

[3] Huang, Y., Chen, Z. X., Tao, Y. U., Huang, X. Z. and Gu, X. F., 2018. Agricultural remote sensing big data: Management and applications. *Journal of Integrative Agriculture*, *17* (9), pp.1915 - 1931. https: //www.sciencedirect. com/science/article/pii/S2095311917618598

[4] Viscarra Rossel, R. A. and Lobsey, C., 2016. Scoping review of proximal soil sensors for grain growing. https: //www.researchgate. net/profile/Raphael - Viscarra - Rossel/publication/317953420_Scoping_review_of_proximal_soil_sensors_for_grain_growing/links/5952ebe2aca272a343db4e27/Scoping - review - of - proximal - soil - sensors - for - grain - growing. pdf

[5] Docs. aws. amazon. com. Visualization Using SageMaker geospatial capabilities - Amazon SageMaker. [online] Available at: https: //docs. aws. amazon. com/sagemaker/latest/dg/geospatial - visualize. html [Accessed 24 Jul.2024].

[6] Toms, S., Crickard, P. and Van Rees, E., 2018. *Mastering geospatial analysis with Python: explore GIS processing and learn to work with GeoDjango, CARTOframes and MapboxGL - Jupyter*. Packt Publishing Ltd. https: //books. google. com/books?hl=en&lr=&id=ivBZDwAAQBAJ&oi=fnd&pg=PP1&dq=The+map+visualization+helps+to+properly+visualize+thp6g

[7] Docs. aws. amazon. com. Amazon SageMaker geospatial capabilities - Amazon SageMaker. [online] Available at: https: //docs. aws. amazon. com/sagemaker/latest/dg/geospatial. html [Accessed 24 Jul.2024].

[8] Newton, T., Villarreal, O. and Verspohl, L., 2017. *Learning D3. js 4 Mapping: Build cutting - edge maps and visualizations with JavaScript*. Packt Publishing Ltd. https: //books. google. com/books?hl=en&lr=&id=lkBPDwAAQBAJ&oi=fnd&pg=PP1&dq=The+map+visualization+helps+to+pro

perly+visualize+the+map+data+and+satellite+images+uploaded+through

[9] Amazon Web Services, Inc. . FAQs | AWS Key Management Service (KMS) | Amazon Web Services (AWS). [online] Available at: https: //aws. amazon. com/kms/faqs/

[10] Lubich, J., Thomas, K. and Engels, D. W., 2019. Identification and classification of poultry eggs: A case study utilizing computer vision and machine learning. *SMU Data Science Review*, *2* (1), p.20. https: //scholar. smu. edu/datasciencereview/vol2/iss1/20/

[11] Janssen, S. J., Porter, C. H., Moore, A. D., Athanasiadis, I. N., Foster, I., Jones, J. W. and Antle, J. M., 2017. Towards a new generation of agricultural system data, models and knowledge products: Information and communication technology. *Agricultural systems*, *155*, pp.200 - 212. https: //www.sciencedirect. com/science/article/pii/S0308521X16305637

[12] Yu, M., Yang, C. and Li, Y., 2018. Big data in natural disaster management: a review. *Geosciences*, *8* (5), p.165. https: //www.mdpi. com/2076 - 3263/8/5/165

[13] Aws. amazon. com. Amazon SageMaker Geospatial Capabilities Now Generally Available with Security Updates and More Use Case Samples | AWS News Blog. [online] Available at: https: //aws. amazon. com/blogs/aws/amazon - sagemaker - geospatial - capabilities - now - generally - available - with - security - updates - and - more - use - case - samples/ [Accessed 24 Jul.2024].

[14] Zhang, Q., Han, L., Jia, J., Song, L. and Wang, J., 2016. Management of drought risk under global warming. *Theoretical and applied climatology*, *125*, pp.187 - 196. https: //link. springer. com/article/10.1007/s00704 - 015 - 1503 - 1

[15] Mauree, D., Naboni, E., Coccolo, S., Perera, A. T. D., Nik, V. M. and Scartezzini, J. L., 2019. A review of assessment methods for the urban environment and its energy sustainability to guarantee climate adaptation of future cities. *Renewable and Sustainable Energy Reviews*, *112*, pp.733 - 746. https: //www.sciencedirect. com/science/article/pii/S1364032119303909