Building Scalable Machine Learning Models on Azure for Predictive Analytics in Business Intelligence

Upesh kumar Rapolu

Houston, USA
Upeshkumar.rapolu@gmail.com

Abstract
The following research project has highlighted the implementation of scalable machine learning models on Azure using predictive analytics in the field of business intelligence. It has explored Azure's cloud services to exemplify its data-driven insights. The combination of Azure Machine Learning and Azure Databricks has paved for allowing successful data ingestion followed by optimum model training and deployment. Moreover, this has understood the role of predictive analytics in business intelligence responsible for building scalable machine learning models. The amalgamation of scalability considerations with collaboration among teams and data governance has been incorporated into the segment of business intelligence.

Keywords: Machine Learning, Azure Machine Learning, Predictive Analytics, Business Intelligence, Azure Databricks, Scalability

I. INTRODUCTION

This research project will underscore the significance of scalable machine learning models on Azure that will be responsible for predictive analytics in business intelligence. The application of scalable machine learning models will stand to be beneficial for accelerating time to value. This will be attained by streamlining prompt engineering and ML model workflows. At the same time, this will be posed to accelerate model development with the powerful infrastructure in business intelligence context. Additionally, the research project will be responsible for observing the tremendous development of data generation that will promote effective resource management. Moreover, the research project will elucidate on understanding the intricate role of predictive analytics in business intelligence and fostering for viewing the Azure services required for machine learning. As a result, this will build scalable machine learning models and will determine best practices for the implementation of predictive analytics on platforms like Azure.

Figure 1: Demonstrating Azure Logo

II. DISCUSSING THE INTEGRATION OF AZURE SERVICES IN MACHINE LEARNING

The following section delves deep by discussing the integration of Azure services in machine learning in a probable manner. It is widely categorised into two parts mainly Azure Machine Learning and Azure Databricks.

Azure Machine Learning: Azure Machine Learning is defined as a systematic process that is used to grant for constructing machine learning solutions which can scale up. At the same time, it is also used to support the customers with the right offers and better services on an overall basis. Fostering a cloud-based environment and nurturing effective tools and services that are considered to be vital in managing machine learning models[1]. This is achieved by getting supported with popular frameworks and tools such as PyTorch and TensorFlow.

Azure Databricks: Azure Databricks is known to be an important parameter which is a unified and open analytics platform. It is variably used to build, deploy and thus maintain enterprise-grade data. At the same time, it also considers AI solutions and analytics at scale. However, the Databricks Data Intelligence Platform tends to integrate with cloud storage and security in the individual cloud account. This seeks to manage and deploy the entire cloud landscape in a cohesive form. Furthermore, the optimisation of the Apache Spark environment provides sustainable results for processing big data and collaboration with machine learning projects.

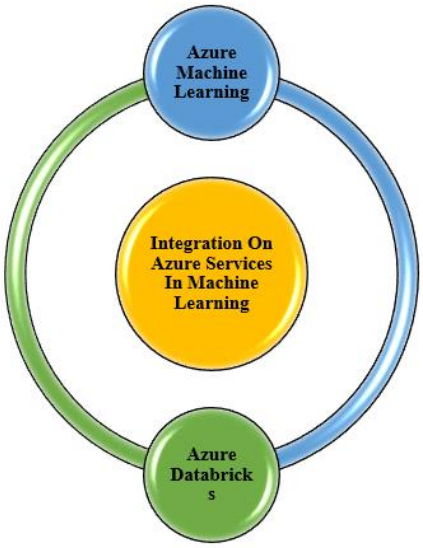

Figure 2: Elaborating Azure Services in Machine Learning

III. UNDERSTANDING THE ROLE OF PREDICTIVE ANALYTICS IN BUSINESS INTELLIGENCE

This section describes the role of predictive analytics in the segment of business intelligence. Predictive analytics plays a vital character in business intelligence in shedding valuable insights into upcoming trends and relative outcomes. At the same time, this aids organisations in conducting ethical decisions and staying ahead of the competition[2]. Predictive analytics stands to be efficient by using statistical algorithms and scalable machine learning approaches to predict the likelihood of future outcomes. However, this is fully

based on the historical data insights that are identified first and then it is analysed within the segment of business intelligence (BI). Moreover, predictive analytics can be used in several domains such as segmenting customer requirements along with forecasting sales and optimisation of supply chain[3]. Having a nuanced understanding of the latest patterns helps organisations to improve their operations and stay at the forefront of competition.

IV. BUILDING SCALABLE MACHINE LEARNING MODELS

Building scalable machine learning models needs to be scalable in mindset by taking into consideration the elements such as dataset size followed by computing resources and model complexity. It also complements with proper preparation of the data. However, this sets the stage by nurturing that scalability optimises the performance of machine learning models. It encompasses several parameters[4]. The forts parameter refers to data ingestion and preparation that can be implemented in a holistic sense to automate data ingestion. It also helps to maintain data integrity and quality. The equation for Data Ingestion is provided below.

V = [D / T],

where V is the ingestion velocity measured in (records per second), D is the total data volume and T is the amount of time taken to ingest the data. The second parameter refers to providing proper training into the model by the utilisation of algorithms in an appropriate form to resolve the problems. It uses AutoML which tends to predict suitable models efficiently[5]. The third parameter is the evaluation and selection of the model. In this case, after the model is trained, then they are evaluated by the implementation of metrics such as precision along with accuracy and recall. The equation supporting model evaluation and selection is described below.

F1 = 2 X [Precision X Recall / Precision + Recall],

where F1 is observed as the ratio of true positive predictions to the actual positives which is intended to highlight the abilities of the models for capturing all the unnecessary instances.

Figure 3: Depicting AutoML

V. ELUCIDATING BEST PRACTICES FOR PREDICTIVE ANALYTICS ON AZURE

The following section illustrates the best practices that are implemented for predictive analytics on Azure. These predictive analytics are mentioned below.

Data Governance: This segment observes that having robust data governance helps to make sure that the quality of the data stays strictly aligned with the regulations. It considers factors such as security measures followed by data lifecycle management data privacy[6].

Scalability Considerations: In terms of scalability considerations, the constructed architecture has the power to get familiarised with the ever-changing workloads to personify to analyse the full potential of Azure[7].

Collaboration Among Teams: Collaborating with the teams helps to allow successful communication with the IT Experts and Data Scientists to replicate sustainable results to facilitate framework. This also helps to adhere to the business choices and technical solutions.

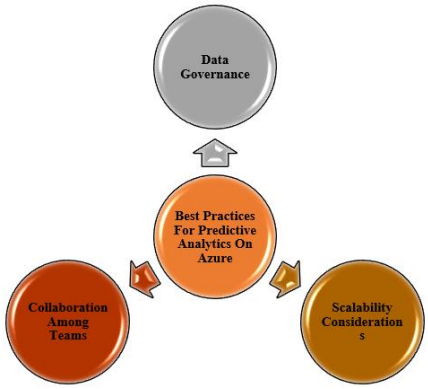

Figure 4: Determining the best practices for Predictive Analytics on Azure

VI. CONCLUSION

The research project has described that the correct integration of scalable machine learning models on Azure for predictive analytics in business intelligence has been observed as a strategic mode. It has helped to uplift the possibilities of justified decisions. The integration of Azure services and tools has enabled organisations to automate data ingestion with model training and deployment. Moreover, this has been augmented by the application of best practices for predictive analytics on Azure such as Data Governance with Scalability Considerations and opting for successful collaboration among the teams. Therefore, this has resulted in boosting the businesses towards constant success in the present competitive infrastructure.

Abbreviations and Acronyms

- AML- Azure Machine Learning
- BI- Business Intelligence
- F1- F1 Score
- ML- Machine Learning
- AI- Artificial Intelligence
- AD- Azure Databricks

Units

- Data Volume is measured in bytes.
- Time is calculated in seconds.
- Ingestion Velocity (V) is measured in records per second

- Accuracy, Precision and FRecall is also calculated in (%)

Equations

- Ingestion Velocity (V) = [D / T],
- F1 = 2 X [Precision X Recall / Precision + Recall], where Precision is the ratio of true positive results to the total predicted positives and recall is the ratio of true positive results to the number of positives.

ACKNOWLEDGEMENT

I hereby acknowledge all the members without whom this research project would not have been completed by me. I would also like to express my sincere gratitude to the professors who have given their valuable time to complete my research project efficiently within the given timeframe.

REFERENCES

[1] A. Botchkarev, "A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms," *Interdisciplinary Journal of Information, Knowledge, and Management,*vol.14,pp.045–076,2019.

[2] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the Carbon Emissions of MachineLearning,"2019.

[3] C. CATAL, K. ECE, B. Arslan, and A. Akbulut, "Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting," *Balkan Journal of Electrical and Computer Engineering*, vol. 7, no. 1, pp. 20–26, Jan. 2019.

[4] J. Duarte *et al.*, "FPGA-Accelerated Machine Learning Inference as a Service for Particle Physics Computing," *Computing and Software for Big Science*, vol. 3, no. 1, Oct. 2019.

[5] V. Derbentsev, N. Datsenko, O. Stepanenko, and V. Bezkorovainyi, "Forecasting cryptocurrency prices time series using machine learning approach," *SHS Web of Conferences*, vol. 65, p. 02001, 2019.

[6] V. E. Pliugin, M. Sukhonos, M. Pan, A. N. Petrenko, and N. Ya. Petrenko, "IMPLEMENTING OF MICROSOFT AZURE MACHINE LEARdNING TECHNOLOGY FOR ELECTRIC MACHINES OPTIMIZATION," *Electrical Engineering & Electromechanics*, vol. 0, no. 1, pp. 23–28, Feb. 2019.

[7] Y. Kumar, S. Kaul, and K. Sood, "Effective Use of the Machine Learning Approaches on Different Clouds,"*SSRNElectronicJournal*,2019.