International Journal of Leading Research Publication (IJLRP)

E-ISSN: 2582-8010 ● Website: www.ijlrp.com ● Email: editor@ijlrp.com

Harnessing AI and Machine Learning for Predictive Analytics in the Sap Data Lake

Naresh Kumar Rapolu

Nareshkumar.rapolu@gmail.com

Abstract

The report has discussed the basic methods of harnessing artificial intelligence and machine learning in order to conduct predictive analytics in the SAP data lake. It has conclusively explained the concepts of artificial intelligence, predictive analysis, machine learning and SAP data management structures. Finally, the report has been concluded with future trends, benefits, challenges and recommendations.

Keywords: SAP Data Lake, Predictive Analysis, Artificial Intelligence, Machine Learning

I. INTRODUCTION

Artificial intelligence is an umbrella term that incorporates multiple technologies to make machines more intelligent and emotionally available. It is a technique to emulate human interaction and intelligence in machine operators. Alternatively, machine learning is a subset of artificial intelligence that focuses exclusively on algorithms. It helps computers to interact with data without explicit programming by pattern identification and data prediction. SAP data lake is basically a repository that has centralised access throughout the SAP ecosystem. It provides accessibility to businesses for storing large data sets. It also has the ability to store structured, semi-structured and unstructured data. Additionally, predictive analysis can help businesses to improve through informed decision-making through data analysis and market forecast. It also helps companies to apprehend associated risks by optimisingoperations and strategizing future patterns. The scope and aim of this research are to understand the importance of artificial intelligence and machine learning for predictive analysis. Further, the assignment will also harness the operations of AI and ML in the SAP data lake to conduct predictive analysis.

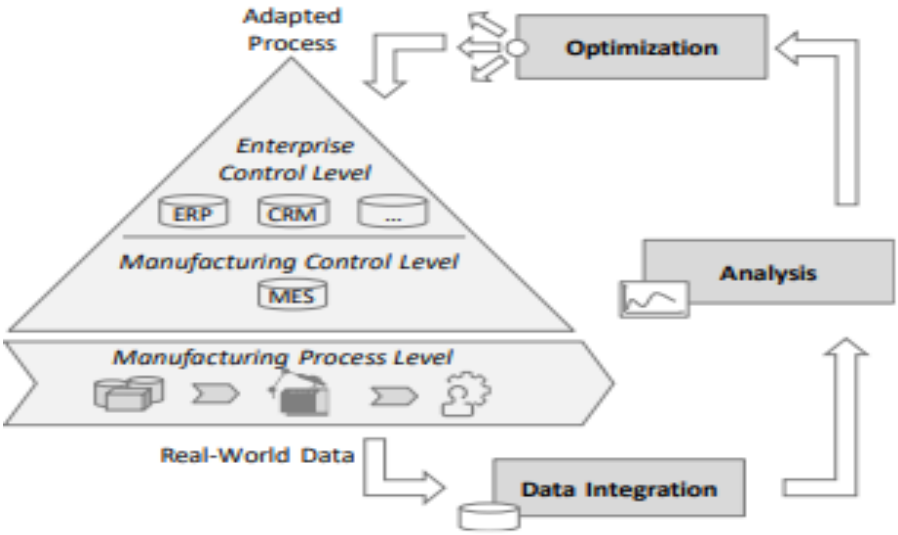

Figure 1: Data-driven manufacturing

II. UNDERSTANDING SAP DATA LAKE

In order to state in a concrete manner data lake is generally a repository of information. It is a type of database that is often confused with data warehouses. However, the usability and accessibility of the lake and warehouse in database management are highly different. This is considering their architecture and demands are different[1]. In case the particularity of a data lake is discussed, it can be stated that it is a cloud setting that is an essential component of contemporary data management strategies. These strategies incorporate the proliferation of social data, the Internet of Things, machine learning, transactional data etc. In the 21st century, 90% of financial institutions rely on Big Data Analytics for future success determination. In 2020 itself, 64.2% zettabytes of digital data were created[2]. Market research and subsequent data analysis predict that by 2026, the estimated value of Data Lake will be $17.6 billion.

III. AI AND MACHINE LEARNING IN PREDICTIVE ANALYTICS

Considering the stability of the supply chain in the contemporary world, manufacturers have been changing their organisational practices towards more alternative methods with the help of predictive analysis. Here it needs to be stated that predictive analysis is the access of statistical algorithms combining internal and external data in order to predict sales and future market trends. The accumulated result of these methods generally helps businesses to optimise their inventory, increase sales, reduce residue cost and improve delivery time[3]. This is the general ideology of predictive analysis; however, when combined with artificial intelligence predictive analysis provides results with more accuracy and authentication.

Considering the COVID-19 pandemic, global business operations at momentarily come to a halt due to the lockdown. During this situation, predictive analysis gained momentum. In the late 2019 and 2020

predictive analysis was used highly by Business leaders to anticipate the economic crisis after the pandemic[4]. It needs to be stated that this was useful and beneficial for organisations. The associated risk analysis proved to help organisations to bounce back from the temporary financial crisis.

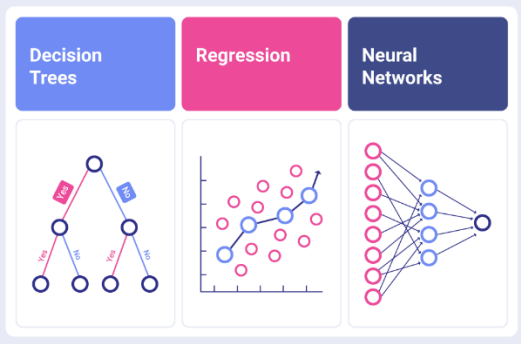

Figure 2: Techniques of predictive analysis

IV. INTEGRATION OF AI/ML IN SAP DATA LAKE

Integration of artificial intelligence and machine learning in SAP data lake provides leverage to businesses regarding predictive model analysis through large amounts of stored data. Identifying patterns is another advantage provided by these integrations because they help to get valuable knowledge regarding the application of machine learning algorithms[5]. The accumulated results of these actions have a direct impact on the unstructured data present in the SAP data lake. In addition to this, it needs to be stated that these types of advanced Data analytics and informed decision-making enrich the structure of the SAP ecosystem.

The fundamental aspects of this integration are provided below:

- ***Data access:*** Data Lake is a centralised repository that helps to transform data without extensive and expensive models. This can be done by the immediate access of artificial intelligence and machine learning models.
- ***Model development:*** Machine Learning algorithms help to use specific tools like SAP data intelligence to customise specific business models[6]. It further helps to get data using artificial intelligence and predictions with machine learning algorithms.
- ***Predictive analytics:*** The ability to analyse patterns in traditional data is helpful for businesses to anticipate future events in the business world. In addition to this it helps to optimise inventory management and understand consumer patterns. Further it also helps to address important issues of provocative cyber threats.

V. BENEFITS AND CHALLENGES

Benefits

- It improves consumer experience and service.
- Helps in resource management through scalability.
- Data integration and cost reduction.

- Scope of future investment.

Challenges

- Unknown domain of operation for multiple groups of employees.
- Inability to provide proper on-site training.
- Risk of inaccuracy.
- Microfocus of AI and ML.

VI. FUTURE TRENDS AND RECOMMENDATIONS

It can be anticipated that artificial intelligence and machine learning will continue to grow and influence business decisions in future[7]. In addition, AI and ML can also be incorporated as cloud-based solutions of database management.

Recommendations

- Data cleaning to remove unidentified bugs.
- Data profiling is based on their structure.
- Holistic transformation of business practices through AI and ML.
- Proper training programs educate internal stakeholders about AI and ML.

VII. CONCLUSION

In conclusion, it can be stated that the integration of artificial intelligence and machine learning in SAP Data Lake needs continuous innovation and improvement. It has been understood that these attributes of modern-daytechnology are successfully helping to revolutionise the business landscape and reduce risk considerably. It needs to be understood that operating artificial intelligence and machine learning applications are somewhat difficult to apprehend. Therefore, it is important to conduct regular checks and balances regarding AI integration in SAP data lake.

Abbreviations and acronyms

- **AI:** Artificial Intelligence
- **ML:** Machine Learning
- **SAP:** Systems, Applications & Products

Units

- **Data Units:** Bytes (B), Kilobytes (KB), Megabytes (MB), Gigabytes (GB), Terabytes (TB), Petabytes (PB)
- **Time Units:** Seconds (s), Milliseconds (ms), Microseconds (μs), Nanoseconds (ns)
- **Computational Units:** FLOPS

Equations

- **Linear regression (used for sales forecasting)**

Single variable: $Y=\beta0+\beta1X+\epsilon$

Multi variable:$Y=\beta0+\beta1X1+\beta2X2+...+\beta nXn+\epsilon$

- **Logistic Regression (used for customer churn prediction, fraud detection, and risk assessment)**

$P(Y=1|X)= 1/1+e^{-(\beta 0+\beta 1X1+...+\beta nXn)1}$

- **Gradient Descent (used for prediction errors)**

$\theta:=\theta-\alpha[\partial J(\theta)/\partial\theta]$

REFERENCES

[1] E. P. Xing , P. Xie , Q. Ho , and D. Wei, "Strategies and Principles of Distributed Machine Learning on Big Data," *Engineering*, vol. 2, no. 2, pp. 179–195, Jun. 2016, doi: https://doi.org/10.1016/J.ENG.2016.02.008

[2] G. Gopal , C. Suter-Crazzolara, L. Toldo , and W. Eberhardt, "Digital transformation in healthcare – architectures of present and future information technologies," *Clinical Chemistry and Laboratory Medicine*, vol. 57, no. 3, Dec. 2018, doi: https://doi.org/10.1515/cclm-2018-0658

[3] K. Beasley, "Council Post: Unlocking The Power Of Predictive Analytics With AI," *Forbes*, Aug. 11, 2021. https://www.forbes.com/councils/forbestechcouncil/2021/08/11/unlocking-the-power-of-predictive-analytics-with-ai/

[4] L. Cao, "Data Science: A Comprehensive Overview," *ACM Computing Surveys*, vol. 50, no. 3, pp. 1–42, Jun. 2017, doi: https://doi.org/10.1145/3076253

[5] M. Algorri, N. S. Cauchon, and M. J. Abernathy, "Transitioning Chemistry, Manufacturing, and Controls Content With a Structured Data Management Solution: Streamlining Regulatory Submissions," *Journal of Pharmaceutical Sciences*, vol. 109, no. 4, pp. 1427–1438, Apr. 2020, doi: https://doi.org/10.1016/j.xphs.2020.01.020

[6] S. A. P. Kumar, R. Madhumathi, P. R. Chelliah, L. Tao, and S. Wang, "A novel digital twin-centric approach for driver intention prediction and traffic congestion avoidance," *Journal of Reliable Intelligent Environments*, vol. 4, pp. 199–209, Oct. 2018, doi: https://doi.org/10.1007/s40860-018-0069-y

[7] S. Lins, K. D. Pandl, H. Teigeler, S. Thiebes, C. Bayer, and A. Sunyaev, "Artificial Intelligence as a Service Classification and Research Directions," *Business and Information Systems Engineering* , vol. 63, no. 4, pp. 441–456, Jul. 2021, doi: https://doi.org/10.1007/s12599-021-00708-w