

Scalable Explainable Ai For Cloud-Based Data Science And Analytics

Omprakash Pandey¹, Dr Sanjay Kumar²

¹Research Scholar Faculty Of Engineering & Technology Asian International University Imphal West, Manipur

²Associate Professor Faculty Of Engineering & Technology Asian International University Imphal West, Manipur

Accepted 20-09-2026

Author(s) Retains the Copyrights of This Article

Abstract

The intersection of Cloud computing and generative artificial intelligence (AI) can be a game changer for business intelligence (BI), particularly when it comes to enhancing predictive analytics capabilities at scale. In this paper, we propose an integrated framework that exploits the elasticity of cloud infrastructure along with the creative problemsolving and data synthesis capabilities of generative AI models. Using generative AI in the cloud empowers organizations to access and apply dynamic data modeling, automated pattern discovery, and real-time forecasting across large, distributed datasets. This study initiation of a scalable architecture for predictive analytics which is costeffective, provides accuracy, and designed to abstract the complexities of the underlying models from business stakeholders enabling smooth business decisions. The model's adaptability across industries with variable volume of data and analytical needs is demonstrated with case studies and simulations. These results reinforce that this integrated point of view greatly enhances performance, agility, and cost-savings in leading-edge BI environments. Therefore, detection of malicious intrusions forms an important part of an integrated approach to network security. In this work, we start by considering the problem of cooperative intrusion detection in WSNs where the nodes are equipped with local detector modules and have to identify the intruder in a distributed fashion. We develop a lightweight ID system, called Lidia, which follows an intelligent agent-based architecture. In Lidia, nodes overhear their neighboring nodes and collaborate with each other in order to successfully detect an intrusion. We show how such a system can be implemented, which components and interfaces are needed, and what is the resulting overhead imposed. We then expand this ID framework with algorithms that incorporate both classes of intrusion detection techniques, i.e., misuse detection and anomaly detection. We investigate in depth some of the most severe routing attacks against sensor networks, namely the sinkhole and wormhole attacks, and we emphasize on strategies that an attacker can follow to successfully launch them. Then we propose novel localized countermeasures that can make legitimate nodes become aware of the threat, while the attack is still taking place. Detailed theoretical analysis and simulation results confirm that the proposed algorithms can always thwart these kinds of attacks. Also, by providing an implementation on real sensor devices, we demonstrate their practicality and efficiency in terms of memory requirements and processing overhead.

Keywords: "Cloud Computing; Generative AI; Predictive Analytics; Business Intelligence; Scalable Architecture; RealTime Forecasting.

Introduction

In this realm some of the most notable architectures are Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs). VAEs allow probabilistic generation of latent representations that are suitable for the tasks of anomaly detection and missing data imputation, whilst GANs are considered analysts of synthetic datasets as they improve the efficiency of predictive models, showing remarkable results in imbalanced classification problems . Previous studies on the use of generative models for data augmentation have been done in the context of

time-series forecasting , fraud detection or customer churn prediction . Used in a cloud ecosystem, these models can scale gracefully in distributed environments and provide real-time inference capabilities. This ensures automated retraining, continuous data ingestion , and pipeline orchestration through a set of platforms including Kubernetes, MLflow and Apache Airflow .Cisco systems (Peter L 2014) estimates that global IP traffic in 2016 will be triple that of 2011 and this exponential growth on Internet usage would limit the quality of service provided to users. Ericsson has predicted the following

demands faced by next generation networks (Ericson Horner 2012) and infers that data rate provided by Long Term Evolution (LTE) will not be enough in the year of 2025

Methodology

The research synopsis must specify how the collected data will be analyzed to answer the research questions/objectives Strategies may include testing of hypotheses, which means that dependent and independent parameters for analysis should be clear as should the type of statistical analysis, the number of observations It is generally recommended that the materials and methods should be written in the past tense, either in active or passive voice. In this section, ethical approval, study dates, number of subjects, groups, evaluation criteria, exclusion criteria

and statistical methods should be described sequentially. The Materials and Methods section briefly describes how you did your research. In other words, what did you do to answer your research question? If there were materials used for the research or materials experimented on your list them in this section. You also describe how you did the research or experiment.

Table of contents

How to write a research methodology.

Why is a methods section important?

Step 1: Explain your methodological approach.

Step 2: Describe your data collection methods.

Step 3: Describe your analysis method.

Step 4: Evaluate and justify the methodological choices you made.

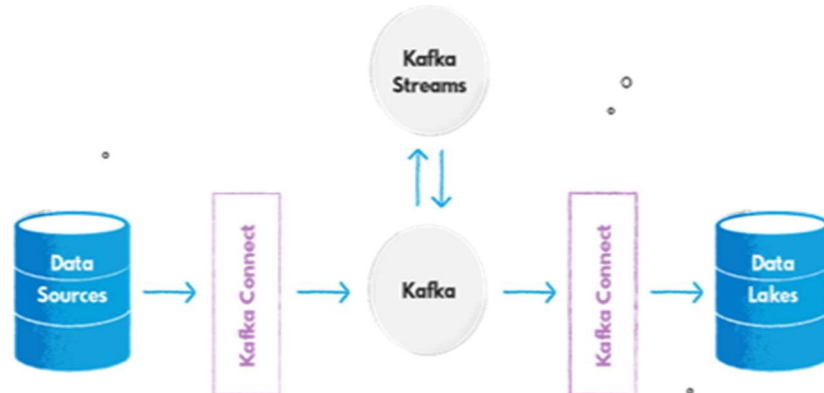


Fig ;1Data Generation Layer

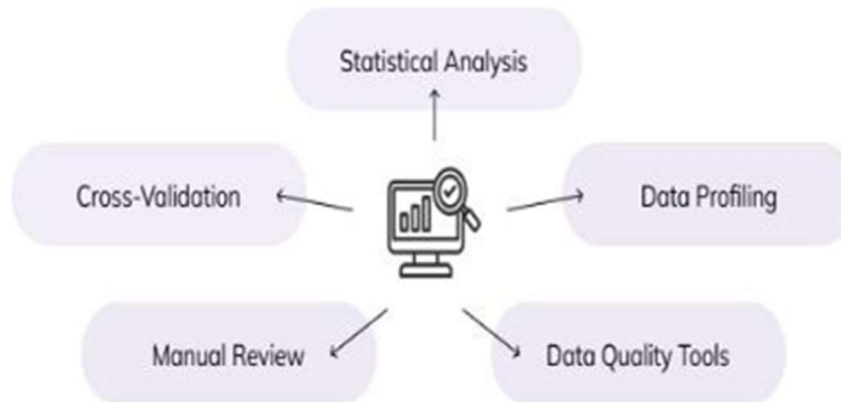


Fig ;2Data Validation Methods

To write a good synopsis, you need to write in the third person; use correct grammar; examine the structure of your novel and include all the main plot points; write

in neutral language; include your hook; ensure you stick to the word count; layer in information about your characters; include all spoilers and plot ... Five

Omprakash Pandey *et. al.*, /International Journal of Engineering & Science Research

examples of research could be surveys, observations, generating research questions, interviews, and focus groups. These examples are dependent on the type of research methodology used. The synopsis These dimensions determine together how well the data represents the phenomena of interest and thus how effectively a predictive model can learn from it. The relationship between data quality and predictive learning is such that predictive learning algorithms, particularly those based on statistical inference or machine learning, rely on the assumption that the training data accurately represents the real-world system from which it was drawn.

Results

Common Sources of Data Quality Issues Data quality problems typically originate from the processes involved in data acquisition, integration, and transformation. Most common sources include: **Bias in Data Collection:** Sampling biases occur when certain sub-populations are under represented, leading to models that generalize poorly to unseen or minority groups. Recognizing the source of quality degradation helps design targeted remediation strategies. Data quality influences every stage of the predictive model lifecycle from data preparation (determines whether training data adequately captures the target distribution), to deployment and monitoring. **Continuous data quality monitoring** is essential to detect drift or degradation that may arise after deployment. **Population:** Population is a well defined set up of all elements pertaining to a given characteristic. It refers to the whole that include all observations or measurements of a given characteristic. Population is also called universe or population. It may be defined as any identifiable and well specified group of individual for example. All primary teachers, no number of all college teachers and all university students are the example of population. A population may be finite or infinite. A finite population is one where all the members can be easily counted. An infinite population is one whose size is unlimited, and cannot count easily. Population of college teachers is an example of finite population and production of wheat, and fishes in river are the example of infinite population.

A measure based upon the entire population is called a parameter. **Sample:** A sample is any number of persons selected to represent the population according to some rule of plan. Thus, a sample is a smaller representation of the population. A measure based upon a sample is known as a statistic. **Sample size:** of selected individual for example, no. of students, families from whom you obtain the require information is called the sample size and usually

denoted by the letter (n). **Introduction to Research Methods in Psychology** Sampling design or strategy: The way researcher selects the sample or students or families etc. is called the sampling design and strategy. It refers to the techniques or procedures the researcher would adopt in selecting some sampling units from which inferences about the population are drawn. **Sampling unit:** Each individual or case that becomes the basis for selecting a sample is called sampling unit or sampling elements. **Quota Sampling** is an improvement over haphazard sampling. It is difficult to represent all population characteristics accurately. Quota sampling ensures that some differences are in the sample. In haphazard sampling, all those interviewed might be of the same age, sex, or background. But, once the quota sampler fixes the categories and number of cases in each category, he or she uses haphazard or convenience sampling. Nothing prevents the researcher from selecting people who act friendly or who want to be interviewed. Quota sampling methods are not appropriate when the interviewers choose who they like (within above criteria) and may therefore select those who are easiest to interview, so, sampling bias can take place. Because not using the random method, it is impossible to estimate the accuracy. Despite these limitations, quota sampling is a popular method among non-probability methods of sampling, because it enables the researcher to introduce a few controls into his research plan and these methods of sampling are more convenient and less costly than many other methods of sampling. **iii) Purposive sampling** Purposive sampling is a valuable kind of sampling for special situations. It is used in exploratory research or in field research. It uses the judgment of an expert in selecting cases or it selects cases with a specific purpose in mind. With purposive sampling, the researcher never knows whether the cases selected represent the population. Purposive sampling is appropriate to select unique cases that are especially informative. For example, a researcher wants to study the temperamental attributes of certain problem behavior children. It is very difficult to list all certain problem behavior children and sample randomly from the list. The researcher uses many different 56 **Introduction to Research Methods in Psychology** methods to identify these cases and approach them to obtain the relevant information

Conclusion

When combining data from multiple sources or data bases (as in multi-data-center systems), discrepancies in formats, timestamps, or identifiers can introduce inconsistencies. **Sensor and Instrumentation Errors:** In Internet of Things (IoT) and industrial systems, hardware malfunction or calibration errors can result

inunreliable measurements. Human Errors: Manual data entry often leads to typographical mistakes, missing fields, or wrong classifications. In modern MLOps (Machine Learning Operations) frameworks, data quality monitoring is as critical as model monitoring. The proponents of this emerging technology envision a future in which environments from nature reserves to cities are instrumented with disposable computing nodes, each with an onboard radio transceiver, battery, environmental sensors and processing capabilities.

Wireless sensor network research grew out of the distributed sensor networks project at the Defense Advanced Projects Research Agency (DARPA), although the technology of the 1970s limited processing and communications and restricted the nodes to large form factors. With the exponential progress and cost reduction in micro processing during the 2000s and 2025s, many new applications for WSN deployment emerged.

Reference

- 1) Fraix-Burnet D. & Devout E. 2015. Stellar populations in ω century: a multivariate analysis, Monthly Notices of the Royal Astronomical Society 450, 3431
- 2) Fraix-Burnet D., Devout E. & Carbone C. 2009. The environment of formation as a second parameter for globular cluster classification, Monthly Notices of the Royal Astronomical Society 398, 1706
- 3) Fraix-Burnet D., Douser E., Choler P. & Overcame A. 2006c. Astrocladistics: a phylogenetic analysis of galaxy evolution II. Formation and diversification of galaxies, Journal of Classification, 23, 57
- 4) Gaskell O. & Steel M. 2006. Neighbor-joining revealed, Molecular Biology and Evolution, 23, 1997
- 5) Harris W. E. 2001. Globular Cluster Systems, Stars-Fields Advanced Course, Vol. 28, Star Clusters, Labhardt L., Bengali B., eds., Springer-Vela Berlin Heidelberg, p. 223