ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

Generative AI-Driven Real-Time Insurance Fraud Detection in Cloud Environments

James Robertson
Asst Professor

Abstract

Rapid advances in Generative AI (GenAI) in recent months open new possibilities for enhancing real-time, online, and incrementally adaptive Machine Learning (ML) models. In this paper, the GenAI paradigm is examined from a cloud feature streaming perspective, enabling a combination of feature engineering and ML model training. Modest inference latency challenges in Online Learning pipeline architectures are resolved using latency-optimized inference pipelines. These concepts are applied to the area of Fraud Detection in Insurance, a phase 1 use case in the ACGA-Cloud Digital Lab, resulting in two contributions: (i) real-time ML models with a 30×39 meilleure separation for Insurance Fraud Detection; and (ii) the application of GANs for transductive domain adaptation and feature augmentation. These contributions are deployed and made available as enhanceML use cases.

Generative AI (GenAI) has recently seen rapid advances, culminating in the release of large natural-language models with unprecedented capabilities. While reflexive initial reactions response in Fine-Tuning, Prompt Engineering, and even Code Generation, the true power of GenAI is revealed with the correct application to traditional modelling domains. In cloud-based ML model training architectures such as Feature Streaming Training and Online Learning, there is plenty of feature engineering required for structural domain knowledge, and at least feature radiating into an ML Model-Space-Super-Space location through within a Generative-Adversarial-Network Combination. The Online Learning pipe is inherently segmented according to subdomain sections of the space, but can be routed with appropriate pipelines to ensure minimal inference latency degradation at these segmentation locations.

Keywords : Generative Artificial Intelligence (GenAI),Real-Time Fraud Detection,Insurance Analytics,Cloud-Based Machine Learning,Deep Learning Architectures,Anomaly Detection Systems,Synthetic Data Generation,Hybrid AI Models,Explainable AI (XAI),Streaming Data Processing,Autoencoder Networks,Model Drift Detection,Scalable AI Infrastructure,Risk Scoring Algorithms,Cyber-Fraud Prevention Systems.

<H1> Introduction

Digital transformation has modernized multiple industries to improve business performance. In the insurance sector, the rapid growth of cloud computing technology and large-scale data management has accelerated the transformation. Generative AI-Enhanced Machine Learning Models are leveraged to improve prediction performance in Insurance fraud detection models. Insurance fraud costs insurers billions each year, spurring many organizations to adopt ML techniques that offer a lucrative opportunity to exploit potential fraud. Radiating features of ML systems are performance indicators. Enhancing the quality of these features enhances overall prediction performance.

While fine-tuning existing models, larger datasets help, but collecting new labeled samples in real time to keep pace with existing model accuracy is impractical. Generative Adversarial Networks under Generative AI technology inject new labeled samples into historical labels of the model. These new samples serve as additional training data for reretaining the fraud detection model before the accuracy

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

decreases. Enabling every ML model in the fraud detection system add such labeled samples when new fraud cases emerge helps to augment prediction performance. Generative AI technology augments the quality of data-injected ML models, enhancing their capability to detect fraud.

<H2> Background and Significance

Generative AI-Enhanced Machine Learning Models for Real-Time Insurance Fraud Detection in Cloud Environments: Hyper-Dimensional Attributes as Key Inputs for Data Fusion and Cross-Validation in Fraudulent-Event Detection. Much research has analyzed the determinants of fraudulent behavior in online and offline contexts. Despite ongoing technological advancements within the insurance sector, insurance frauds continue to be the primary concern of the insurance industry and practitioners. Traditional machine learning (ML)-based approaches have established a benchmark for performance, but the methods are now becoming outdated primarily due to three factors. Real-time inference and operationalization are becoming more pressingly relevant, as fraud detection is turning into a real-time function. Latency and execution cost are becoming the key contributors during inference. Generative AI techniques—such as Generative Adversarial Network (GAN), Denoising Diffusion Probabilistic Model (DDPM), and text generation—are now becoming the trend, with companies investing heavily in these areas.

Real-time data ingestion and feature engineering have paved the way for hyper-dimensional conceptualization. Hyper Dimensional Computing (HDC) is a new neural computing paradigm inspired by biological systems. HDC explores random patterns—high-dimensional and redundant binary or bipolar vectors—and develops applications in cognitive-related areas like recognition and retrieval. HDC views deep neural networks (DNNs) as special-purpose learning machines supported by an extremely large number of micro-modules that are inherently noise-resilient. Self-organizing neural networks take feature permutation and combination to the next level for noise-resilient talking and cross-validation.

Table 1: Core Components of the Proposed Architecture

Stage	Component	Description	Role in Fraud Detection
1	Data Ingestion	Real-time streaming via Kafka	Collects live transactional & monitoring data
2	Generative Flow	Feature radiating pipeline	Enhances raw features using GenAI
3	Feature Streaming	Asynchronous stream architecture	Enables real-time feature updates
4	GAN-based Augmentation	Conditional GANs	Handles class imbalance
5	ML/DL Modeling	Hybrid models (CML, DNN)	Fraud classification
6	Latency-Optimized Inference	Serverless optimized pipelines	Reduces response time
7	Operationalization	Cloud deployment	Scalable fraud detection

<H1> Theoretical Foundations

Generative AI methods differentiate from conventional machine-learning algorithms, including different categories of artificial neural networks and other machine-learning methods. Generative-adversarial networks and variational autoencoders are the two main techniques in generative AIs. The two networks of generative-adversarial networks have game relationships between themselves. The first network generates pattern data, while the second network attempts to identify the generated pattern data according to real data distribution. The primary purpose of generative-adversarial networks is to make these two networks perform better.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

Variational-encoder uses a latent variable model whose distribution can be made close enough to the real data-generating distribution after training. Classification performance will be improved if patterns undergo generative methods before entering classification networks.

An insurance-fraud-detection architectural design divides the entire treatment procedure into several processing sections: (1) data ingestion, (2) generative-flow for feature-radiating data processing, (3) asynchronous-set feature-stream architecture, (4) generative-ai-enhanced modeling techniques, (5) real-time inference, and (6) experimental methodology. Data-ingestion section responsibility includes stream processing of the data source by enhancing consequential properties. Generative-flow design concentrates on using data-generating flow for focusing a relevant flow of it into a detach and additional processing path parallel to its original processing flow. A class-spread-gated design structure allows data particles belonging to different fraud classes to enter the reserved generative-flow section.

<H2> Generative AI and Its Role in Radiating Features

Generative AI has emerged as a paradigm that exemplifies the ease with which systems can be trained to generate a type of data (e.g., images, videos, textual content) that closely resembles the original training data and can radiate features that impact multiple other data analysis tasks. Large generative models have been trained to serve as feature generators for various downstream applications at scale, such as OpenAI's CLIP model for image–text similarity. In these models, the generative aspect is of secondary importance; the primary purpose is to enhance the performance of related models by learning well-distributed features in the latent space. For instance, CLIP has been leveraged to improve cross-modal retrieval tasks with fewer annotations than direct training.

Theoretically similar approaches can be devised for tabular data by employing Generative Adversarial Networks (GANs). These models mathematically devise special latent spaces where it is much easier to sample from the subspaces of data classes. Such latent spaces can be used for data augmentation to enable models to learn well-regionalized–distributed decision boundaries. These models thereby radiate robust features for separate but related tasks. While direct training in an end-to-end fashion with a task-specific architecture is always an option, such an approach generally requires significantly more data than in the radiate-and-reuse paradigm.

<H1> Architectural Considerations in Cloud Environments

Data ingestion and feature streaming represent fundamental challenges when machine learning models are deployed in a cloud environment. While data are usually stored in a blob storage dedicated to batch inference, the models must support the ingestion of small batches in real time in order to achieve the desired latency. Requests load a subset of the features required for inference while maintaining the overall model performance. Moreover, the predictions must be performed with minimal latency. The latency can be reduced not only by using a suitable cloud service (e.g., an inference optimization solution) but also by simplifying the inference pipelines of the models.

Due to the need for a minimal time between feature engineering and prediction, the TS-ML proposal is extended with data streaming. A dedicated service enables the storage of the data in a stream managed by Apache Kafka. Generative AI models can be devised that provide a small number of features needed to make inference with a specific set of features, without meaningfully degrading the performance.

<H2> Data Ingestion and Feature Streaming

Within a cloud environment supporting continuous operation, a data ingestion layer relays constant streams of online events for model inference. Streaming data are sourced from transactional applications and from monitoring systems such as third-party fraud detection services,

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

adapting a tactical-scoring approach. Redundant payload information, noise, and superfluous data elements extraneous to model inference are removed—unlike batch-oriented data-mining effort, where all factors are considered important. For each prior transaction, intelligent systems provide a set of generated features contoured through reinforcement-learning techniques. Instead of waiting for new labeled data, data providers gradually extend the labeled space to model feature- and label-efficient machine-learning hierarchies for a chain of models.

Online-transaction-processing source systems register events in approximately-real-time while maintaining near-zero latencies. To predict business transactions in online-transaction-processing systems such as eCommerce or other platforms where a transactional process occurs, intelligent-tactical-scoring models minimize material loss caused by high-risk transactions. Combating security challenges in cloud computing is a necessity. Security-related roles are provided in the security domain and by security service providers dedicated to monitoring the cloud infrastructure around-the-clock and providing real-time alerts and suggestions. Information generated by such monitoring-services should be availed for operation as additional features while monitoring such systems. Redundant information across such systems can be effectively characterized through available correlation, resulting in additional features derived through reinforcement learning models.

<H1> Generative AI-Enhanced Modeling Techniques

Generative models, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and diffusion models, play a pivotal role in data-driven machine learning. Thanks to these models, data generation has become significantly easier. However, Generative AI can also simplify the creation of other kinds of intelligent systems. A naive view of the problem can lead to thinking only about data-driven modeling techniques and not about those systems in which all features of a complex problem can be easily radiated and streamed in real time to an ML or DL algorithm for online inferencing. Such a system must have its main core connected through low-latency streams to routing engines that feed all the main generated events to dedicated ML, DL, or hybrid inference engines. The events must then traverse the routing engines, and most of them composed of a minimum number of features must traverse the low-latency pipelines dedicated to the traditional ML approach. The other events containing additional features but still not enough to start inference engines must traverse the intermediate-latency engines. This is when latency must be weighted with the increase in the number of features ready for complementing the signature of the samples sent to the inference engines.

Although the integration of Generative AI techniques seems to be bypassed, it can support the Data Ingestion layer by providing suitable services to generate synthetic data. Data generation is crucial when working with a low amount of samples in a supervised context. For instance, some classes can be strongly imbalanced with very few samples to learn the corresponding model. In such cases, Generative Adversarial Networks (GANs)—or support from any other Generative AI model that can provide convincing samples—can be beneficial. Even in this context, with the generation of new plausible samples, one of the main concerns is that the data remain realistic, and the model must not learn to distinguish real from fake samples.

Table 2: Generative AI Techniques Used

Technique	Purpose	Benefit
GAN	Data augmentation	Improves minority-class learning
Conditional GAN	Tabular fraud data synthesis	Reduces imbalance
VAE	Latent distribution modeling	Better feature representation

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

Technique	Purpose	Benefit
Diffusion Models	Tabular modeling	High-quality synthetic samples
Reinforcement Learning	Feature optimization	Long-term utility maximization

<H2> Generative Adversarial Networks for Data Augmentation

Ample labeled training data is essential for designing generative AI-based supervised learning models with good predictive performance. However, in many real-world applications, data may be either scarce or highly imbalanced. Data augmentation strategies—techniques for creating new training data from existing training data—have become a popular solution approach for these challenges. GANs offer a family of data augmentation techniques that can be applied for a wide variety of data types, including images of many modalities, video, time series, and tabular data. GANs use two deep learning models: a generator and a discriminator. The generator's goal is to create data that resembles real-world data, while the discriminator's goal is to differentiate between real-world data and the data generated by the generator. Training of the generator and the discriminator occurs in an adversarial manner, resulting in a generator that generates realistic data.

GANs can be tailored to augment tabular datasets originating from traditional data sources, such as databases and spreadsheets. Conditional GANs leverage both labeled and unlabeled data and provide realistic data augmentation for imbalanced tabular datasets. Data from a minority class commodity is used to train a deep learning model that identifies characteristic features of that commodity. A conditional GAN is constructed based on these features and trained using the minority class data together with a moderately large subset of the majority class data. After training, the GAN generates fictional records of the minority class commodity, which are then incorporated into the training dataset for the original deep learning model in order to reduce prediction error for that class.

<H1> Real-Time Inference and Operationalization

Operationalizing CML models for real-time inference requires a comprehension of the latency demands of the business applications, projected model updates, and variations in system load during the day. Such considerations may lead to a set of decision trees or a multi-task CML model for a subset of the labels that requires particularly low-latency predictions in the real world. For the remaining labels, the models might be hosted as dynamic web services, with demand responding to the system load or by calling sequence prediction techniques. When the modelling of all labels does require very low latency, CML modelling techniques may be combined with the previous temporal fusion technique to provide predictions at infra-second intervals.

In temporal fusion, the inference pipeline for a single vanishing label is moved within global routing manager and the continuously updating representations for the other labels are instead fed into a cached predictive/decision system that generates one prediction for each label at regular intervals. Latency-optimized inference pipelines and CML models are not orthogonal and can in fact be combined, generating a new output for the non-CML labels only when either the routing manager or the decision trees for those labels ask for a new prediction. In this case, load balancing concepts from network engineering can be applied to the randomization of the future execution of the CML pipelines.

<H2> Latency-Optimized Inference Pipelines

Lower latency systems require the use of latency-sensitive inference pipelines optimized for execution across edge-cloud and server-less environments. Supported by an automated data generation service defined in an earlier section, the core processes in the deployed inference

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

pipelines are implemented in server-less mode, thereby allowing the service provider to economically respond to bursts in traffic. Authentication and fraud detection services ingest data generated by the service and this is streamed into the Authentication and Anti-Fraud Models designed for low latency. For fraud detection, spam attack prevention, and Online Analytical Processing systems, outgoing responses for closed-loop services routed across a logical edge service. The Authentication Models deploy three separate ML Models, but with a single optimized service, thereby achieving a high level of confidence in predictions. The speed of processing is improved by integrating a light-weight Image Processing Service, dedicated to improving image quality, so as to reduce the time taken by the Identification Model in extracting people from group images.

Performance decay is evident when low-quality images, i.e., images with blurriness, distance, illumination, and facial occlusion degrading their quality, are fed into the model. To mitigate such performance degradation issues, an online light-weight Image Processing Service is designed. The motivation is to enhance the input images by correcting distortions, artefacts, and interruptions in real-time. The processing sequence includes de-blurring, de-illuminating, and face detection methods, even though a complete restoration to the original high-resolution image quality is not guaranteed. However, the marked enhancement in quality facilitates better face extraction and contributes positively to the model's accuracy.

<H1> Evaluation Methodologies

Objective investigations demonstrate the efficacy of radiating features via Generative AI, data ingestion through stream processing, Generative Adversarial Networks for data augmentation, latency-optimized inference pipelines, and operationalization in a simulated cloud environment by exposing the complete detection workflow on-dedicated boards. Metrics evaluating model performance and overall overhead across model operation converge in late evening or weekend hours to medium or low values replicating normal production traffic. Robust OpenFlow forwarding builds secure connections and avoids packet overhead rendering a compact implementation possible. Benchmark datasets containing multiple features or a mix of quantitative/qualitative features alongside TensorFlow are deployed to expose the various concepts on cloud-supported multitenant infrastructures with the possibility of carrying out scheduled checks by exposing parts of the functionality on-dedicated boards.

Recent processes or attack features proposed or exploited recently are different from the traditional detection concepts used in the cloud environment; a generative model based on Generative Adversarial Networks is suggested for augmenting datasets; data ingestion is managed through stream processors such as Apache Kafka designed for continuous lossless data patterns; feature radiating exploration is facilitated by Generative AI; operational phases from data ingestion up to classification checks are packed into pipelines separately deployable on-demand in their dedicated infrastructure; reinforcement learning with a Deep-Q-Network maximizes long-term utility. The ability to predict loss of data or association and prepare for preventive/mitigative action is essential for near-real-time detection in the financial domain rather than direct action during attack occurrence.

<H2> Benchmark Datasets and Simulation Environments

Leading cloud service providers support the use of data augmentation techniques with device-optimized accelerator boards (GPU, TPU) in the backend. However, the inference phase latency is critical not just for customers but also for the service providers themselves. Therefore, it is expected that the cost of using augmentation techniques will keep decreasing and that service providers will also use data-augmentation techniques to reduce the actual model inference time. Insurance datasets suitable for real-data augmentation have shown to run efficiently by directing the trained model with optimally chosen operational regionizations with margin scores and thereby supporting function-based manipulation and synthetic data creation.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

These models not only consume lesser time in inference-supporting cloud environments but also bring down the actual active latency of the models supporting several concurrent parallel calls.

Various benchmark datasets are available for training supervised classifier frameworks for cyber gaming, insurance-detection, face-recognition, and other applications with generative and replay memory networks. The operating/simulation environment of the model plays a crucial role in inference speed and overall size and capacity of the trained model rather than the simulated device and model architecture. Genetically customized environments with margin score resemblance on distributed groups demand lesser-time-consuming model inference and deliver a real-time inference-supporting latent space. The condition of keeping the augmented variations under control for synthetic-data generation is essential to maintain the validation and test accuracy. Furthermore, wide augmentations in the operational regionization built by segmentation-based or other direct repartitioning are preferred.

<H1> Conclusion

The emergence of large-scale generative models capable of data generation in a multi-modal setup represents a step-function advance for the broader Machine Learning community. Generative AI extends beyond just training data supply and provides a multitude of opportunities to radiate features for the downstream Machine Learning components in cloud-based environments. Cloud-based architecture for any Machine Learning application relies on subsystems optimally realized for dedicated objectives. A short description of these aspects follows. The trend of providing features as a service will continue to gain traction. Feature Engineering as a Service, Feature Store as a Service, Real-Time Feature Streaming, and Feature Sourcing from other Machine Learning Models are expected to gain momentum.

The emergence of generative models capable of enabling data augmentation for machine learning applications supports the definition of Generative Adversarial Networks (GAN) as a Service and GAN-based data generation for supervised learning task training. A cognitive capability that enables systems to respond accurately to new input cannot be defined as true artificial intelligence. A feature of intelligence is the ability to expect uncertainty. Risk forecasting models enabling such a capability are often built for customer claims in the insurance sector. Machine Learning models built for this purpose cannot process new inflow claims in real time without an appropriate inference pipeline around them. An objective-based low-latency inference pipeline, especially tailored for this requirement, is one of the goals.

<H2> Emerging Trends

With the vast increase of both the usability of Generative AI technologies, especially their capacity to create high-fidelity data, as well as the negative impacts of malicious use, the emersing capability of Generative AI to radiaite and augment features in data sets is gaining traction. Such approaches, which include both image augmentation, feature enhancement in video datasets, as well as natural language processing tasks in supervised machine learning approaches, have been recognized as a promising, forward-looking trend. The ability of Generative AI to support specialized feature engineering and additionally augment real-world footage collected, while presenting a refined set of features for supervised machine learning models, has also been taken into account by the insurance industry, where image-data-based policy fraud is a growing concern. Moreover, it is believed that such approaches can be used to encourage higher-class business development – the refinement and introduction of a new class of detected business acceleration, reputation, and credibility augmentation fraud-based use cases for existing companies.

Emerging within the insurance field, the emersing technology in real-time insurance fraud detection in cloud environments has been recognized as one of the aviation-related business functions posing a higher risk of policy categories in relation to the growing level of crime and fraudulent use of new technologies and services. The

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

introduction of Generative Adversarial Network (GAN) algorithms to uplift, forecast, and augment real-life datasets for e.g. machine–learning-based computer-use-fingerprint detection schemes has clearly illustrated the scalability of such approaches. In aviation industry-agnostic domains, the successful introduction of fraud-related detection in the service level have also been demonstrated, where Generative AI-assisted GAN datasets-related augmentation schemes have been applied in presence of low to mid-size classes with ML models capable of flagging and marketing their detection capability having the user being the key player in detection.

<H1> References

1. Gomes, C., Jin, Z., & Yang, H. (2021). Insurance fraud detection with unsupervised deep learning. Journal of Risk and Insurance, 88(3), 591–624.
2. Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences, 557, 317–331.
3. Pandugula, C., Ganti, V. K. A. T., & Mallesham, G. (2024). Predictive Modeling in Assessing the Efficacy of Precision Medicine Protocols. EDUCATIONAL ADMINISTRATION: THEORY AND PRACTICE Учредители: Green Publication БИБЛИОМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ: Входит в РИНЦ: на рассмотрении Цитирований в РИНЦ: 0 Входит в ядро РИНЦ: нет Цитирований из ядра РИНЦ: 0 Рецензии: нет данных Процентиль журнала в рейтинге SI: ТЕМАТИЧЕСКИЕ НАПРАВЛЕНИЯ:.
4. Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. IEEE Access, 9, 165286–165294.
5. Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. Expert Systems with Applications, 193, 116429.
6. Hassan, H. B., Barakat, S. A., & Sarhan, Q. I. (2021). Survey on serverless computing. Journal of Cloud Computing, 10, 39.
7. Polineni, T. N. S., Kumar, A. S., Maguluri, K. K., Koli, V., Valiki, D., & Ravikanth, S. (2024, November). A Scalable and Robust Framework for Advanced Semi Supervised Learning Supporting Universal Applications. In Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024).
8. Błaszczyński, J., de Almeida Filho, A. T., Matuszyk, A., Szeląg, M., & Słowiński, R. (2021). Auto loan fraud detection using dominance-based rough set approach versus machine learning methods. Expert Systems with Applications, 163, 113740.
9. Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. Research in International Business and Finance, 62, 101744.
10. Singireddy, S., Adusupalli, B., Pamisetty, A., Mashetty, S., & Kaulwar, P. K. (2024). Redefining financial risk strategies: The integration of smart automation, secure access systems, and predictive intelligence in insurance, lending, and asset management. Journal of Artificial Intelligence and Big Data Disciplines, 1(1), 109-124.
11. Abe, K., Kim, S., Nojima, R., Ozawa, S., & Moriai, S. (2022). Privacy-preserving federated learning for detecting fraudulent financial transactions in Japanese banks. Journal of Information Processing, 30, 789–795.
12. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730–27744.
13. Bhowmik, A., Sannigrahi, M., Chowdhury, D., Dwivedi, A. D., & Mukkamala, R. R. (2022). DBNex: Deep belief network and explainable AI based financial fraud detection. In Proceedings of the 2022 IEEE

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

International Conference on Big Data (pp. 3033–3042). IEEE.

14. Shyamala Anto Mary, P., Kalisetty, S., & Mandala, V. M. (2024, November). Advancing IoT Data Forecasting with Deep Learning Framework for Resilience Scalability and Real-World Applications. In Proceedings of the 3rd International Conference on Optimization Techniques in the Field of Engineering (ICOFE-2024).
15. Shafiei, H., Khonsari, A., & Mousavi, P. (2022). Serverless computing: A survey of opportunities, challenges, and applications. ACM Computing Surveys, 54(11s), Article 239, 1–32.
16. Marin, E., Perino, D., & Di Pietro, R. (2022). Serverless computing: A security perspective. Journal of Cloud Computing, 11, 69.
17.
18. Recharla, M. (2024). Antioxidants, Biological Markers, Catalase, Glutathione Peroxidase, Chronic Periodontitis, Saliva, Smokeless tobacco, Smoker. Frontiers in Health Informatics, 13(8), 4999.
19. Li, Z., Guo, L., Cheng, J., Chen, Q., He, B., & Guo, M. (2022). The serverless computing survey: A technical primer for design architecture. ACM Computing Surveys, 54(10s), Article 220, 1–34.
20. Xiao, S., Bai, T., Cui, X., Wu, B., Meng, X., & Wang, B. (2023). A graph-based contrastive learning framework for Medicare insurance fraud detection. Frontiers of Computer Science, 17(2), 172341.
21. Mashetty, S. (2024). Research insights into the intersection of mortgage analytics, community investment, and affordable housing policy. Available at SSRN 5249213.
22. Debener, J., Heinke, V., & Kriebel, J. (2023). Detecting insurance fraud using supervised and unsupervised machine learning. Journal of Risk and Insurance, 90(3), 743–775.
23. Lu, J., Lin, K., Chen, R., et al. (2023). Health insurance fraud detection by using an attributed heterogeneous information network with a hierarchical attention mechanism. BMC Medical Informatics and Decision Making, 23, 62.
24. Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decision Analytics Journal, 6, 100163.
25. Pamisetty, A. (2024). Leveraging Big Data Engineering for Predictive Analytics in Wholesale Product Logistics. Available at SSRN 5231473.
26. Fanai, H., & Abbasimehr, H. (2023). A novel combined approach based on deep autoencoder and deep classifiers for credit card fraud detection. Expert Systems with Applications, 217, 119562.
27. Chen, Z.-Y., & Han, D. (2023). Detecting corporate financial fraud via two-stage mapping in joint temporal and financial feature domain. Expert Systems with Applications, 217, 119559.
28. Yi, Z., Cao, X., Pu, X., Wu, Y., Chen, Z., Khan, A. T., Francis, A., & Li, S. (2023). Fraud detection in capital markets: A novel machine learning approach. Expert Systems with Applications, 231, 120760.
29. Nandan, B. P. (2024). Semiconductor Process Innovation: Leveraging Big Data for Real-Time Decision-Making. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 4038-4053.
30. Zhou, Y., Li, H., Xiao, Z., & Qiu, J. (2023). A user-centered explainable artificial intelligence approach for financial fraud detection. Finance Research Letters, 58, 104309.
31. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.
32. Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized LLMs. Advances in Neural Information Processing Systems, 36.
33. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In International Conference on Learning Representations.

ISSN: 3067-4166

AMERICAN DATA SCIENCE JOURNAL FOR ADVANCED COMPUTATIONS

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 22 REVISED: NOVEMBER 19 ACCEPTED: NOVEMBER 27 PUBLISHED: DECEMBER 17

34. Mashetty, S. (2024). The role of US patents and trademarks in advancing mortgage financing technologies. European Advanced Journal for Science & Engineering (EAJSE)-p-ISSN, 3050-9696.
35. Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. (2024). Self-RAG: Learning to retrieve, generate, and critique through self-reflection. In International Conference on Learning Representations.
36. Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 12, 157–173.
37. OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., et al. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774.
38. Kreuzberger, D., Kühl, N., & Hirschl, S. (2023). Machine learning operations (MLOps): Overview, definition, and architecture. IEEE Access, 11, 31866–31879.
39. Pamisetty, V. (2024). Transforming taxation systems through predictive analytics and AI-driven compliance monitoring tools. Am Data Sci J Adv Comput, 3, 55-68.
40. Schrijver, G., Sarmah, D. K., & El-hajj, M. (2024). Automobile insurance fraud detection using data mining: A systematic literature review. Intelligent Systems with Applications, 21, 200340.
41. Banulescu-Radu, D., & Yankol-Schalck, M. (2024). Practical guideline to efficiently detect insurance fraud in the era of machine learning: A household insurance case. Journal of Risk and Insurance, 91(4), 867–913.
42. Hong, B., Lu, P., & Yang, Y. (2024). Health insurance fraud detection based on multi-channel heterogeneous graph structure learning. Heliyon, 10(9), e30045.
43. Vorobyev, I. (2024). Fraud risk assessment in car insurance using claims graph features in machine learning. Expert Systems with Applications, 251, 124109.
44. Motie, S., & Raahemi, B. (2024). Financial fraud detection using graph neural networks: A systematic review. Expert Systems with Applications, 240, 122156.
45. Paleti, S. Agentic AI in Financial Decision-Making: Enhancing Customer Risk Profiling. Predictive Loan Approvals, and Automated Treasury Management in Modern Banking.
46. Awosika, T., Shukla, R. M., & Pranggono, B. (2024). Transparency and privacy: The role of explainable AI and federated learning in financial fraud detection. IEEE Access, 12, 64551–64560.
47. Abdul Salam, M., Fouad, K. M., Elbably, D. L., & Elsayed, S. M. (2024). Federated learning model for credit card fraud detection with data balancing techniques. Neural Computing and Applications, 36, 6231–6256.