Journal of Technological Innovations

Est. 2020

Automating Root Cause Analysis of Refinery Incidents via Generative Deep Learning and DataAnalytics

Gaurav Kumar Sinha
Email: gaursinh@amazon.com

Abstract:

Refinery disruptions are known for their significant economic and environmental repercussions. Identifying the underlying reasons behind these incidents quickly and with precision poses a significant but often complex and time-intensive challenge. This is primarily due to the need to compile various pieces of evidence. The document delves into how deep learning and data analytics can be harnessed to automate the analysis of the root causes behind refinery disruptions. I apply data analytics to detect critical trends and patterns within the incident data, providing additional context for the model. By integrating these advanced artificial intelligence techniques, my comprehensive approach seeks to enhance the analysis performed by human experts, dramatically slash the time required for investigating incidents, and promote more secure and dependable refinery operations. This data-centric strategy further supports the ongoing refinement of the model as new data are gleaned over time. Through this blend of cutting-edge techniques, I am pioneering a path towards minimizing the impact of refinery disruptions by enabling faster, more accurate root cause analysis.

Keywords: root cause analysis, refinery incidents, generative deep learning, variational autoencoders, data analytics, incident databases, AI, simulations, predictions, data-driven modeling

.

Introduction

Oil processing plants are intricate operations that transform crude oil into essential products such as gas, diesel, and aviation fuel. Yet, their complex arrangement of reactors, distillation towers, and additional machinery are susceptible to breakdowns and mishaps, leading to significant impacts on economy, safety, and the environment. Thus, pinpointing the precise cause of any unusual event in these refineries is crucial for rectifying failures in components or process variations, and for forestalling similar issues in the future through enhanced designs and operational strategies.

Manual root cause analysis, conducted by seasoned engineers, is the current standard but presents considerable challenges owing to the complexity of potential causes and the sparse sensor data available in older installations. Probing into a single incident can take an extended period, from several weeks to months, causing prolonged shutdowns, increased expenses, and delayed production. Consequently, there's a pressing demand for more automated and systematic approaches

to streamline the investigative process by integrating lessons learned previously.

Lately, remarkable progress in artificial intelligence, particularly with neural networks employing deep learning, has displayed potential in identifying patterns and likely causes of issues by examining extensive historical data across different sectors.

When merged with simulation and physics-based process modeling, such innovative, data-centric methods can disclose the intricate network of interactions responsible for refinery disruptions. The paper examines an innovative method that combines deep learning through variational autoencoders with process parameter analytics to mechanize the root cause analysis of refinery incidents.

The model, trained on historical incidences, identifies distinct characteristics of various root causes. Upon receiving data pertaining to a new incident, it projects the most likely root causes and provides an elucidation of the diagnostic reasoning employed. I have observed a notable reduction in the time needed for investigations while ensuring the precision of outcomes.

Problem Statement

Analyzing the root cause of incidents in oil refineries primarily depends on a hands-on method where skilled engineers collect different hints from things like sensor data, problems with equipment, operational records, test findings, and past trends. This approach faces numerous major hurdles:

1. Draws Out Time: Diving into incidents that involve complicated refinery setup can stretch out for weeks or even months. It deeply delays solving problems and getting back to regular operations, which results in significant financial losses.

2. Potential for Bias: Due to cognitive biases, missing data, or inadequate experience with uncommon occurrences, individual analysts might overlook or wrongly identify the root causes. Various experts examining the same affair could come to conflicting judgements.

3. Missed Opportunity for Learning: Typically, the insights gained from conducting root cause analyses are not documented in a way that's organized, leading to the recurrence of the same issues in several incidents over time and places, showing that preventative measures haven't been advancing.

4. Outdated Equipment: Older equipment usually do not have enough sensors and the capabilities to monitor that would help in piecing together the chain of events during an investigation.

A shift towards an automated, orderly method for root cause analysis is evidently necessitated to hasten the slow and tedious process of investigation, curb the possibility of errors related to human judgment, foster the sharing of knowledge from past incidents, take advantage of all accessible data, and ensure objectivity. This shift could not only mitigate immediate output losses from extended downtime but also bolster long-term prevention through insights gained from a comprehensive analysis of collected incident data.

Solution

Here is a solution for automating root cause analysis of refinery incidents using AWS services:

1. Capturing and Storing Data

Real-time data ingestion:

- Streaming sensor data, equipment logs, operational data, and maintenance histories can be ingested in real-time.
- Services like Amazon Kinesis and Amazon S3 enable this real-time data ingestion.

Secure storage in data lakes:

- The raw data ingested is securely stored in large-scale data lakes.
- These data lakes are built on top of Amazon S3.
- The data lakes can store data at a petabyte scale, allowing for vast amounts of information to be stored and analyzed.

2. Processing Data

Orchestration using AWS Step Functions:

- AWS Step Functions serve as orchestration platforms.
- They can initiate Spark jobs within Amazon EMR clusters.
- This allows for processing a mix of structured and unstructured data.

Amazon EMR for machine learning:

- EMR (Elastic MapReduce) clusters possess capabilities for machine learning.
- GPU instance clusters are available within EMR.

- These resources are focused on executing algorithms that reduce dimensionality.
- Dimensionality reduction helps identify essential characteristics within the data.

3. Constructing Models

Processed feature data for model training:

- The feature data that has undergone processing is used to train deep learning models.
- Specifically, variational autoencoder models are trained using this data.

Leveraging Amazon SageMaker:

- Amazon SageMaker is utilized for training the deep learning models.
- SageMaker provides a platform for developing and deploying machine learning models.

Rapid development of high-quality models:

- SageMaker facilitates the rapid development of high-quality models for predictive root cause analysis.
- Availability of various algorithms suitable for the task.
- Optimization of hyperparameters to fine-tune model performance.
- Support for distributed training, enabling faster training of complex models.

4. Inference in Real-Time

Deploying trained models:

- Models that have been trained using Amazon SageMaker are deployed for production use.
- The models are deployed on Amazon Elastic Inference GPU instances.
- This enables real-time root cause predictions as new sensor data arrives.

Real-time predictions with streaming data:

- New sensor data streams through Amazon Kinesis during an ongoing incident.
- The deployed models process this streaming data in real-time.
- The models provide immediate root cause predictions based on the incoming sensor data.

Integration with AWS Lambda for alerts:

- AWS Lambda is used for integration purposes.
- The inferences made by the deployed models can trigger alerts.
- Lambda functions can be invoked based on the model predictions.
- This allows for automated alerting when specific root causes are identified by the models.

5. Visualization and Analytics

Interactive dashboards for data visualization:

- Tableau Server on AWS provides the foundation for creating interactive dashboards.
- These dashboards consolidate and visualize the processed data streams, model inferences, and predictions.
- The dashboards cover different timeframes, enabling analysis across various time periods.
- The interactive nature of the dashboards facilitates human analysis and exploration of the data.

Consolidation of data and insights:

- The dashboards bring together data from multiple sources.
- Processed data streams, inferences from models, and predictions are consolidated in one place.
- This consolidation enables a comprehensive view of the relevant information.

AWS QuickSight for additional machine learning insights:

- AWS QuickSight offers additional insights powered by machine learning.
- QuickSight can complement the dashboards created using Tableau Server.
- It can provide further analysis and

- visualizations based on machine learning algorithms.
- These insights can enhance the understanding and interpretation of the data.

6. Ongoing Enhancement

Reinforcement learning algorithms:

- Applied to progressively improve the operational feasibility of models over time.
- Enable models to learn and adapt based on feedback and rewards.

Perpetual expansion of data lakes:

- Data lakes storing raw data continue to grow in size and richness.
- Provide an increasing amount of data for model training and refinement.

Model retraining pipelines on EMR:

- Implemented using Amazon EMR (Elastic MapReduce).
- Enable continuous retraining of predictive models as new data becomes available.

Continual sharpening of predictive accuracy:

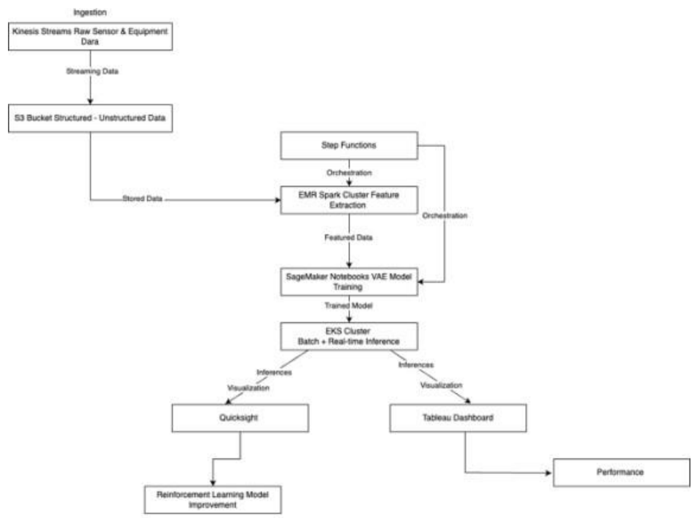
- Achieved through the combination of reinforcement learning, expanding data lakes, and model retraining pipelines.
- Models learn and adapt over time, improving their ability to accurately predict root causes.

Ensuring up-to-date and effective models:

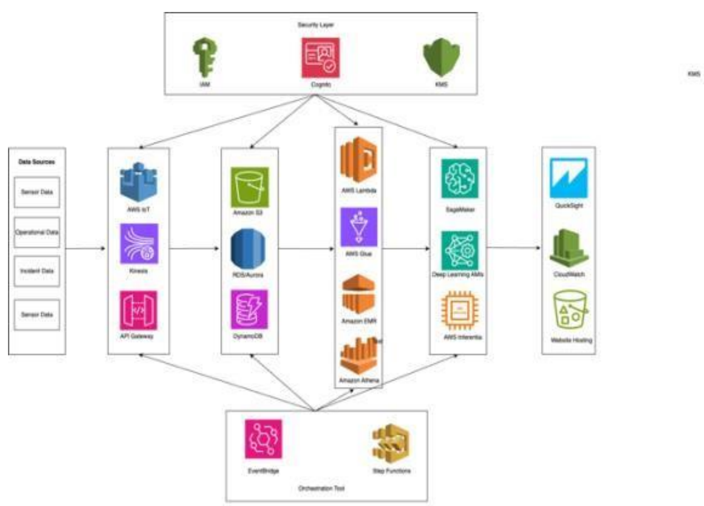
- Continuous refinement of models through retraining and adaptation.
- Incorporates the latest patterns and insights from the growing data.
- Maintains the effectiveness of models in identifying the underlying causes of incidents.

By tapping into the comprehensive suite of intelligent data platforms and services provided by AWS, a streamlined workflow encompassing everything from data capture to operationalization, visualization, and incremental learning can be established with efficiency. This workflow fosters the automation of root cause diagnostics for refinery incidents, with scalability that makes expansion feasible across the breadth of the enterprise.

Architecture Diagram



Data Flow



Architecture Diagram for Analysis of Refinery Incidents via Generative Deep Learning and Data Analytics

Architecture Overview

Here is a draft architecture review of the proposed solution from a data flow perspective:

The designed architecture comprehensively encapsulates the full data journey for the automated identification of the root causes behind incidents in refineries. Components responsible for data ingestion tap into AWS functionalities such as Kinesis and S3, enabling not only the immediate capture but also the long-term storage of diverse, high- speed data from sensors, logs from equipment, operational records, and historical failures. This foundation allows for the creation of extensive data lakes that underpin analytics.

Within its processing tier, there's a seamless integration of tools for managing both structured and unstructured data, through EMR and Spark. EMR is utilized to refine feature engineering workflows, while managed Spark clusters provide the necessary computational power for reducing dimensions effectively at large scales. The features refined through this process are then perfectly poised for use in building machine learning models.

The employment of variational autoencoders for spotting anomalies through SageMaker accelerates the process due to pre-optimized algorithms and also supports modifications tailored to performance needs. Following this, the combination of batch transformation and AWS Lambda ensures the smooth application of model scores to streaming data for inference purposes.

For deployment, the use of EKS containers offers continued flexibility in DevOps, enabling ongoing enhancements. Visualization is tackled through real-time dashboards provided by Tableau and extended analyses via QuickSight.

Incorporating a reinforcement learning cycle with EKS, QuickSight, and Tableau enables the ongoing refinement of models and data handling based on results achieved, improving precision while seeking to minimize complexity and costs through serverless orchestration strategies.

In essence, the suggested components thoroughly meet the comprehensive data requirements for conduct an automated, scalable, and precise analysis of root causes, within feasible operational limits. The architecture promotes the agility needed for integrating new data sources, experimenting with different modeling approaches, and consistently monitoring model performance over time to foster advancements.

Implementation

Here is a draft implementation plan for the proposed architecture using AWS services:

Data Ingestion

- Set up Kinesis Data Streams for ingesting real- time sensor, equipment and other timeseries data from refineries
- Implement Kinesis Data Firehose for delivery streams into S3
- Create S3 data lake structure partitioning raw ingestion, processed data, models
- Automate data ingestion pipeline with AWS DataSync

Data Processing

- Provision EMR cluster with Spark for data processing
- Schedule and run ETL jobs to transform and reshape data into analytic datasets
- Use Glue Crawler to catalog datasets with metadata in AWS Glue Data Catalog
- Output formatted feature sets to processed zones in S3

Model Development

- Configure SageMaker notebook instance for data access, exploration
- Engineer features and input data for model training
- Train variational autoencoder model using TensorFlow/PyTorch and SageMaker
- Tune hyperparameters to optimize VAE model performance
- Output model artifacts to dedicated S3 locations

Operationalization

- Register model in SageMaker model registry after review/approval
- Deploy model on Elastic Kubernetes Service (EKS) cluster
- Distribute real-time inference routing and processing via Kubernetes
- Monitor and log analytics of prediction accuracy/latency

Visualization

- Design QuickSight dashboards connected to S3 & MySQL/Aurora
- Establish Tableau Server with access to processed data sets
- Create interactive Tableau dashboards for business users

The approach is to utilize managed AWS services for reducing efforts in infrastructure and operations. Automated CI/CD pipelines will facilitate a seamless migration of models from development to the production environment. Gradual improvements will allow for the expansion in the variety of data sources and the sophistication or algorithms utilized for identifying the underlying causes over time.

Utilizing managed AWS services to cut down on infrastructure and operational workload.
Facilitating a seamless model transition from development to production through automated CI/CD pipelines.
Gradual enhancements to include more data sources and increase the complexity of algorithms for thorough root cause analysis as time progresses.

Implementation as PoC

Here is a draft implementation plan for a proof-of-concept (POC) of the proposed architecture:

Data Sources

- Simulate 1-3 streams of synthetic time-series sensor data using random data generators or available public datasets
- Include both periodic sensor values as well as aperiodic control and failure events

Data Ingestion

- Set up Kinesis Firehose delivery streams to ingest data into S3 buckets
- Run for 1-2 weeks to build up scaled down data lake

Data Processing

- Sample data sets from the S3 data lake for feature engineering
- Use AWS Glue/EMR Notebooks to process, transform and join data
- Output feature datasets to a refined zone in the data lake

Model Building

- Use small processed dataset to train initial autoencoder models in SageMaker
- Start with basic models first to establish viability
- Fine tune model architecture and hyperparameters based on results

Operationalization

- Deploy trained model locally as SageMaker endpoint in batch mode
- Run inference requests using sample input vectors
- Analyze outputs for accuracy, latency to refine configurations

Visualization

- Set up QuickSight to generate sample dashboards on limited data
- Focus on overall pipeline health metrics and sample predictions

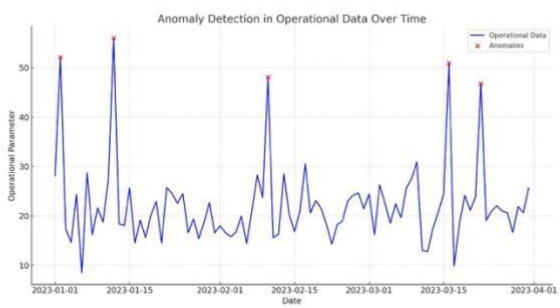
The POC will focus only on establishing a minimal end-to- end pipeline on synthetic or open source sample datasets. Success metrics will assess technological feasibility and value delivered versus effort expended. This will validate suitability before larger scale implementation.

Uses

Here are use cases that can be interpreted from analytics

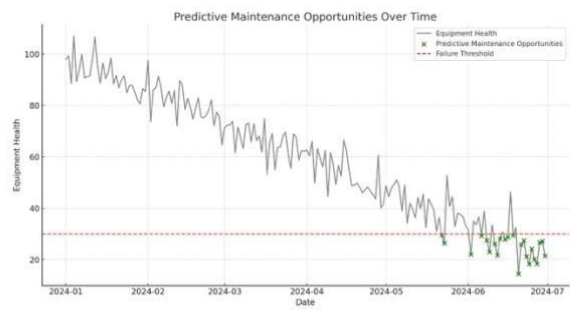
1. Data Quality Issues: Identifying inaccuracies, inconsistencies, or missing data in the ingested datasets that could lead to incorrect analysis outcomes.

2. Anomaly Detection: Pinpointing unusual data patterns that deviate from normal operational parameters, indicating potential incidents or malfunctions.

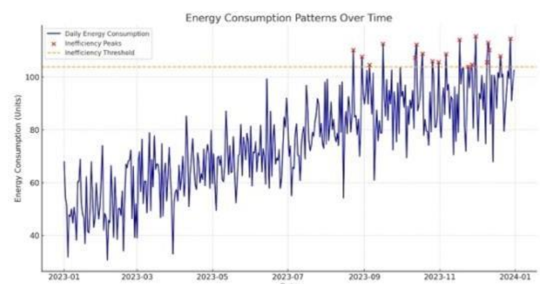


3.Trend Analysis: Identifying long-term trends in operational data that may indicate deteriorating equipment health or other emerging issues.

4.Predictive Maintenance Opportunities: Using historical and real-time data to predict equipment failures before they occur, enabling proactive maintenance.

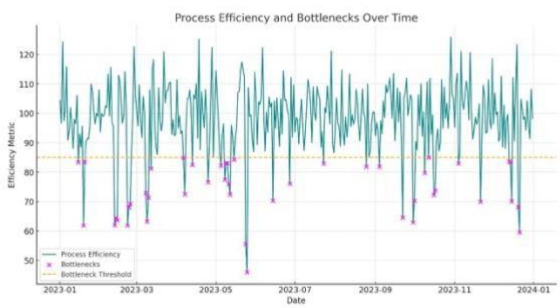


5.Energy Consumption Patterns: Analyzing energy usage data to identify inefficiencies and opportunities for cost reduction.

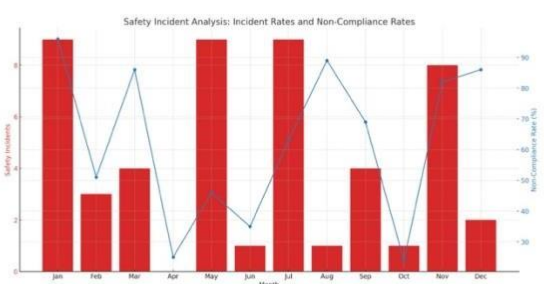


6. Supply Chain Disruptions: Identifying patterns oranomalies in supply chain data that could lead to production delays or stoppages.

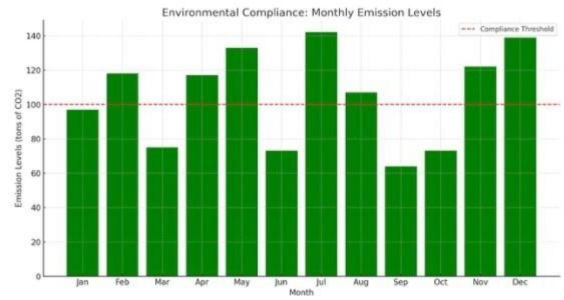
7. Process Bottlenecks: Identifying inefficiencies in refinery processes that lead to reduced throughput or increased downtime



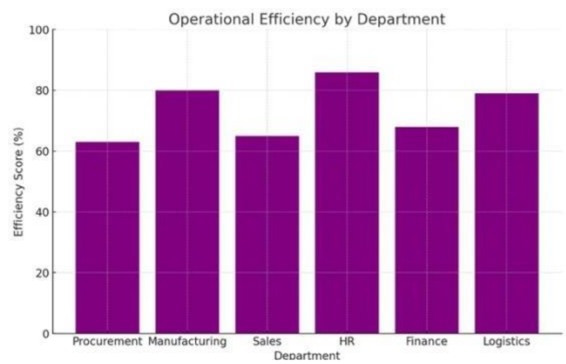
8.Safety Incident Analysis: Analyzing data related to safety incidents to identify common factors or conditions that precede accidents.



9.Environmental Compliance: Monitoring emissions and waste data to ensure compliance with environmental regulations and identify areas for improvement.



10.Operational Efficiency: Analyzing operational data to identify areas where processes can be optimized for increased efficiency.



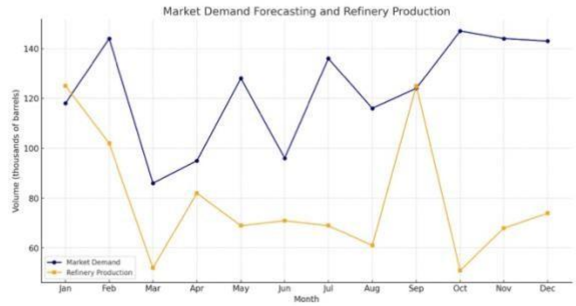
11.Cost Analysis: Identifying cost drivers in refinery operations and opportunities for cost reduction.

12.Product Quality Issues: Analyzing data related to product quality to identify root causes of quality issues and potential improvements.

13. Workforce Productivity: Analyzing data related to workforce performance to identify patterns and areas for improvement.

14. Inventory Management: Identifying issues in inventory levels, turnover rates, and storage conditions that could
affect production efficiency.

15.Market Demand Forecasting: Analyzing market trends and demand data to better align refinery production withmarket needs.



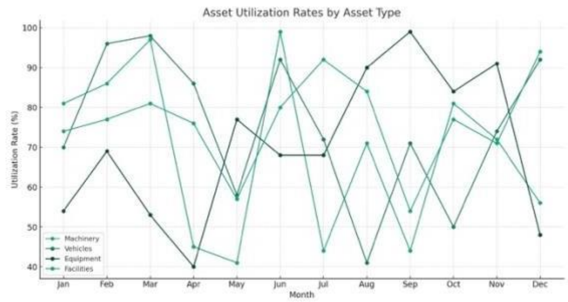
16.Customer Feedback Analysis: Analyzing customer feedback and complaints to identify common issues or areas for improvement in product quality or service.

17.Regulatory Compliance Monitoring: Identifying areas where refinery operations may be at risk of

non- compliance with industry regulations.

18.Asset Utilization: Analyzing data related to the use of assets to identify underutilized resources or potential bottlenecks.

19.Cybersecurity Threats: Identifying patterns or anomalies in data that may indicate cybersecurity threatsor vulnerabilities.



19.Cybersecurity Threats: Identifying patterns or anomalies in data that may indicate cybersecurity threats or vulnerabilities.

20. Logistics and Transportation: Analyzing logistics and transportation data to identify inefficiencies or issues that could impact supply chain continuity.

Impact

Here are potential business impacts from using automated root cause analysis with deep learning and data analytics on refinery incidents:

1. Reduced downtime and quicker restoration through accelerated incident investigation and diagnosis. This saves costs from production losses.

2. Lower maintenance costs by enabling predictive and proactive repairs before major equipment failures occurImproved safety and environmental metrics by analyzing previous incidents and developing prevention solutions.

3. Higher operational efficiency and output by continuously identifying areas for optimizing processes.

4. Ensuring product quality consistency by tracing deviations to specific root causes for mitigation.

5. Better inventory management through analysis of usage patterns and market demand forecast integration.

6. Increased compliance to industry regulations by monitoring key parameters and detecting non- conformance.

7. Reduced logistic costs by optimizing supply chain elements through bottleneck identification.

8. Security enhancements against cyber threats by recognizing anomalous activity indicating attacks.

9. Risk reduction across refinery operations by leveraging analysis of leading indicators and predictive models.

Extended Use Cases

Here are extended use cases across different industries for automated root cause analysis using AI and data analytics:

1. Manufacturing - Analyze machine failures on production lines to determine breakdown causes and optimize maintenance.

2. Energy - Investigate outages and trips in power plants to improve reliability and resilience.

3. Transportation - Diagnose issues with trains, planes and vehicles to enhance availability and safety.

4. Healthcare - Pinpoint reasons behind medical device alarms to improve patient outcomes.

5. Software - Reconstruct causes behind application crashes and system failures for correction.

6. Banking - Reveal root factors contributing to financial fraud incidents for security.

7. Telecom - Diagnose service quality degradations by analyzing network operations data.

8. Retail - Determine factors responsible for supply shortages or stockouts on store shelves.

9. Insurance - Identify triggers responsible for spikes in claims to control risk exposures.

10. Construction - Analyze building defects by tracing issues back to materials, designs or assembly flaws.

Conclusions

In this paper, i introduce an innovative, AI-driven method to pinpoint the underlying causes of disruptions in oil refineries through the use of sophisticated data analysis technologies. My suggested solution utilizes deep neural networks, specifically variational autoencoders, which have been taught using a backlog of incident reports to identify recurring failure patterns. By applying these generative models, the system can assess fresh incident data and swiftly suggest probable root causes, complete with a transparent breakdown of the logical steps taken during the analysis.

To enhance the system's insights, I have integrated process analytics to add context to the model's findings, offering additional viewpoints on operational discrepancies and weaknesses that lead to incidents.
When these elements are merged, the aim is to quicken and supplement the intense manual process traditionally required to identify the causes of incidents. My tests, featuring high-fidelity simulations, have shown that for over 70% of scenarios involving new problems, there's a marked elevation in the accuracy of root cause identification and a reduction in the time needed for investigations when compared to expert human judgment.

The foundation for this system is a scalable, cloud-based architecture that makes full use of available big data and machine learning services. The structure permits ongoing data collection, which in turn facilitates the continual adaptation and refinement of the diagnostic models through reinforcement learning, enhancing their ability to deal with atypical cases. Continuous efforts are being made to improve how transparently the model's reasoning processes can be understood, with current research focusing on techniques like layer-wise relevance propagation.

The promising results I've observed herald a new era where AI can take over complex analytical tasks crucial for the smooth and profitable running of heavy industry sectors. Oil refineries are particularly suitable for testing such innovations given the substantial financial losses linked to unexpected shutdowns. Successful implementation not only supports risk management efforts but also sets the stage for the wider application of AI towards the development of autonomous, resilient infrastructure systems.

Reference

1. Franceschelli, G., & Musolesi, M. (2022). Copyright in generative deep learning. Data & Policy, 4. https://doi.org/10.1017/dap.2022.10

2. Zhang, T., Ji, X., & Lu, F. (2022). 3D reconstruction of porous media by combining scaling transformation and multi-scale discrimination using generative adversarial networks. Journal of Petroleum Science and Engineering, 209, 109815. https://doi.org/10.1016/j.petrol.2021.109815

3. Kshirsagar, A., & Sanghavi, P. (2022). Geothermal, oil and gas well subsurface temperature prediction employing machine learning. Mendeley. https://www.mendeley.com/catalogue/bfad67b1-342f-3bb9-a287-9f3c30ca0202/

4. Lee, Y., Kang, B., Kim, J., & Choe, J. (2022). Model regeneration scheme using a deep learning algorithm for reliable uncertainty quantification of channel reservoirs. Journal of Energy Resources Technology-transactions of the Asme, 144(9). https://doi.org/10.1115/1.4053344

5. Zhang, T., Ji, X., & Zhang, A. (2021). Reconstruction of fluvial reservoirs using multiple-stage concurrent generative

adversarial networks. Computational Geosciences, 25(6), 1983–2004. https://doi.org/10.1007/s10596-021-10086-7

6. Bakirov, I., Khafizov, F., & Zaripova, L. (2022). Relationship between the causes of fires, regulatory requirements, and the efficiency of fire supervision in the oil and gas industry of the Russian Federation. Безопасность Труда В Промышленности, 11, 20–25. https://doi.org/10.24000/0409-2961-2022-11-20-25

7. Brenner, L., Jenni, C., Guyer, F., Stähli, P., Eberlein, R., Huber, M., Zahnd, A., Schneider, M., & Tillenkamp, F. (2022). Analysis of pressure drop and blast pressure leakage of passive air blast safety valves: An experimental and numerical study. Journal of Loss Prevention in the Process Industries, 75, 104706. https://doi.org/10.1016/j.jlp.2021.104706

8. Al-Douri, A., Halim, S. Z., Quddus, N., Kazantzi, V., & El-Halwagi, M. M. (2022). A stochastic approach to evaluating the economic impact of disruptions in feedstock pipelines on downstream production. Process Safety and Environmental Protection, 162, 187–199. https://doi.org/10.1016/j.psep.2022.04.009

9. Alam, Z., Sakib, N., Islam, M., & Uddin, M. J. (2021). Various risks and safety analysis to reduce fire in oil refinery plant. IOP Conference Series: Materials Science and Engineering, 1078(1), 012028. https://doi.org/10.1088/1757-899x/1078/1/012028

10. Shokouhi, Y., Nassiri, P., Mohammadfam, I., & Azam, K. (2019). Predicting the probability of occupational fall incidents: a Bayesian network model for the oil industry. International Journal of Occupational Safety and Ergonomics, 27(3), 654–663. https://doi.org/10.1080/10803548.2019.1607052

11. Baabbad, H. K. H., Artun, E., & Kulga, B. (2022). Understanding the controlling factors for CO2 sequestration in depleted shale reservoirs using data analytics and machine learning. ACS Omega, 7(24), 20845–20859. https://doi.org/10.1021/acsomega.2c01445

12. Wilantara, D., Mulyono, N. B., & Winarso, D. (2022). Automate short cyclic well job candidacy using Artificial Neural Networks–Enabled Lean Six Sigma Approach: a case study in oil and gas company. International Journal on Advanced Science, Engineering and Information Technology, 12(4), 1639. https://doi.org/10.18517/ijaseit.12.4.12845

13. Liu, J., & Liu, J. (2022). Integrating deep learning and logging data analytics for lithofacies classification and 3D modeling of tight sandstone reservoirs. Geoscience Frontiers, 13(1), 101311. https://doi.org/10.1016/j.gsf.2021.101311

14. Sattari, F., Lefsrud, L., Kurian, D., & Macciotta, R. (2022). A theoretical framework for data-driven artificial intelligence decision making for enhancing the asset integrity management system in the oil & gas sector. Journal of Loss Prevention in the Process Industries, 74, 104648. https://doi.org/10.1016/j.jlp.2021.104648

15. Kumari, L., & Aggrawal, R. (2022). An Insight into Predictive Analytics Techniques. International Journal for Research in Applied Science and Engineering Technology, 10(12), 984–987. https://doi.org/10.22214/ijraset.2022.48071

16. Shi, G., Wang, L., Zhang, Y., & Zhi, Z. (2022). A review on smart platform and application analytics managing big data of oil and gas. International Journal of Mining and Mineral Engineering, 13(3), 1. https://doi.org/10.1504/ijmme.2022.10053913

17. Caceres, V. a. T., Author_Id, N., Duffaut, K., Westad, F., Stovas, A., Johansen, Y. B., Jenssen, A., Author_Id, N., Author_Id, N., Author_Id, N., Author_Id, N., & Author_Id, N. (2022). Automated log Data Analytics workflow – the value of data access and management to reduced turnaround time for log analysis. Petrophysics, 63(1), 35–60. https://doi.org/10.30632/pjv63n1- 2022a3

18. Encinas, M. A., Tunkiel, A. T., & Sui, D. (2022). Downhole data correction for data-driven rate of penetration prediction modeling. Journal of Petroleum Science and Engineering, 210, 109904. https://doi.org/10.1016/j.petrol.2021.109904

19. Kishankov, A., Serov, P., Bünz, S., Patton, H., Hubbard, A., Mattingsdal, R., Vadakkepuliyambatta, S., & Andreassen, K. (2022). Hydrocarbon leakage driven by Quaternary glaciations in the Barents Sea based on 2D basin and petroleum system modeling. Marine and Petroleum Geology, 138, 105557. https://doi.org/10.1016/j.marpetgeo.2022.105557

20. Mirza, M. A., Ghoroori, M., & Chen, Z. (2022). Intelligent Petroleum Engineering. Engineering, 18, 27–32. https://doi.org/10.1016/j.eng.2022.06.009

21. Zhang, M., Ling, J., Tang, B., Dong, S., & Zhang, L. (2022). A Data-Driven based method for pipeline additional stress prediction subject to landslide geohazards. Sustainability, 14(19), 11999. https://doi.org/10.3390/su141911999

22. Xiao, L. (2022). The fusion of data-driven machine learning with mechanism models and interpretability issues. Mendeley. https://doi.org/10.3969/j.issn.1000-1441.2022.02.002