Enhancing Data Security and Data Sensitivity: Leveraging the Synergy of Blockchain & Artificial Intelligence

By Vikas Jain & Sachin Chouhan

ABSTRACT

1. Challenges in the Digital Era:

- Data Sensitivity Concerns: Cyber threats, breaches, and information exposure pose significant challenges.
- Increasing importance of protecting sensitive user data.

2. Synergy of Technologies:

- Blockchain's Secure Foundation: Decentralized, immutable ledger for secure authentication and data storage.
- Use scenarios demonstrate protection of medical records, financial information, and personal identifiers.
- AI's Dynamic Role: Threat detection, anomaly analysis, and user authentication for enhanced data security.
- Blocks unauthorized access to confidential information.

3. Comprehensive Data Security Architecture:

- Synergy of Blockchain, AI, and ML: Intelligent data protection through machine learning.
- Real-world examples showcase effective collaboration against modern threats.

Introduction

Data Challenges:

- Rising importance of data in enterprises.
- Increasing difficulties in security and sensitivity.

Paradigm Shift:

- Vulnerabilities in traditional data protection methods.
- Necessity for transformative approaches.

Blockchain and AI Synergy:

- Blockchain for decentralized, tamper-proof data integrity.
- AI's proactive defense with advanced threat detection.

Motivation

Avoiding a Cautionary Tale: Policy Considerations for Artificial Intelligence in Health Care

By Jodi G. Daniel & Aaron Cummings on November 17, 2023

POSTED IN ARTIFICIAL INTELLIGENCE

On November 8, 2023, the Senate Health, Education, Labor and Pensions (HELP) Committee Subcommittee on Primary Health and Retirement Security discussed the impact of artificial intelligence (AI) on the healthcare sector in the Committee's second AI hearing in nine days. The hearing comes as the White House and Congressional leaders seek to quickly respond to AI threats, mitigate its dangers, and harness its potential for American industry. Senators discussed the recent Executive Order issued by the White House to guide AI regulation and innovation across all sectors, including in the health and human services sectors.

Tackling the terrors of insurance fraud with AI

According to the Insurance Fraud Detection Market Research, 203, the global insurance fraud detection market size was valued at $3.3 billion in 2021 and is projected to reach $28.1 billion by 2031, growing at a CAGR of 24.2% from 2022 to 2031.

Introduction

ML Integration:

- ML enhances AI for intelligent data classification and encryption.
- Creation of a robust data security framework.

Research Objective:

- Explore integration of blockchain, AI, and ML.
- Fortify data security and protect sensitive information.

Data Security Challenges in AI Era:	**Blockchain's Role in Enhancing Data Security:**	**Decentralized AI Model Training and Challenges:**
The widespread adoption of AI demands access to vast sensitive	Blockchain's inherent properties, such as decentralization and	The integration of AI and blockchain introduces decentralized
data, posing concerns about data breaches and unauthorized	immutability, are recognized for providing a tamper-resistant	approaches to AI model training, mitigating concerns about
access. Researchers explore integrating AI and blockchain as a	and distributed ledger. Anonymization techniques like data	centralized data breaches. However, challenges such as
solution for robust data protection.	hashing and homomorphic encryption are explored to protect	scalability, regulatory compliance, and ethical considerations
	sensitive data during AI model training.	must be addressed to ensure responsible and secure data usage.

Methodology

1. Data Extraction from Blockchain using API

- Developed API for targeted data extraction from Ethereum blockchain.
- Focused on timestamps and block rewards.
- Ensured privacy by excluding personally identifiable information.

2. Anonymization of Sensitive Data

- Applied advanced anonymization techniques (e.g., data hashing, differential privacy).
- Safeguarded user identities before storing data on the blockchain.

Methodology

3. Random Forest Regression Model

- Utilized anonymized dataset for training.
- Implemented Random Forest ensemble learning for predicting block rewards.
- Accommodated uneven and non-linear data distributions.

4. Model Evaluation and Validation

- Reserved a portion of the dataset for testing.
- Evaluated model accuracy using metrics like MSE or RMSE.
- Employed cross-validation techniques for robustness validation.

Methodology

5. Deployment and Prediction

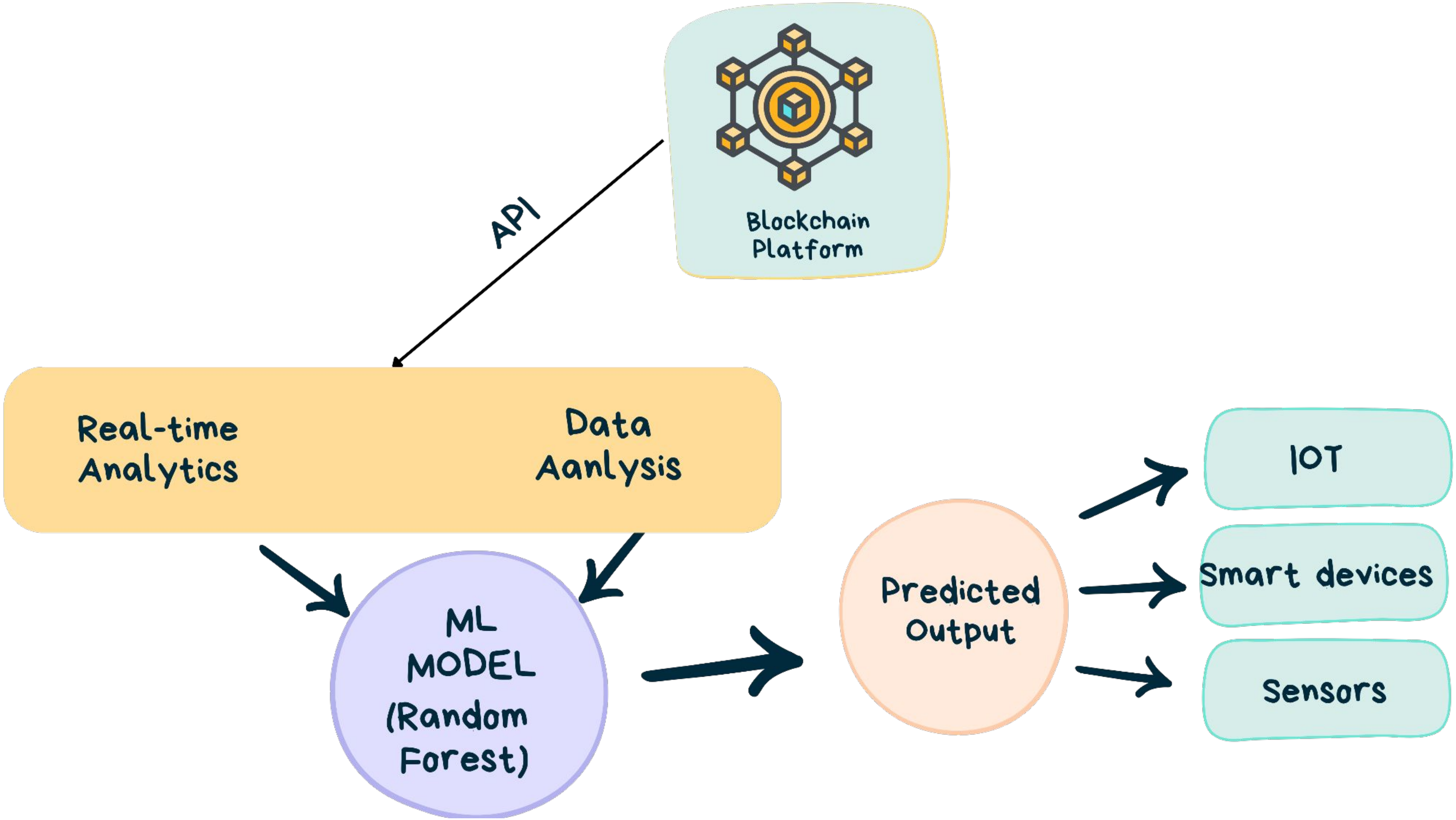
- Deployed validated model for real-time block reward predictions.
- Enhanced transparency and decision-making within the blockchain network.

6. Applicability to Various Blockchains

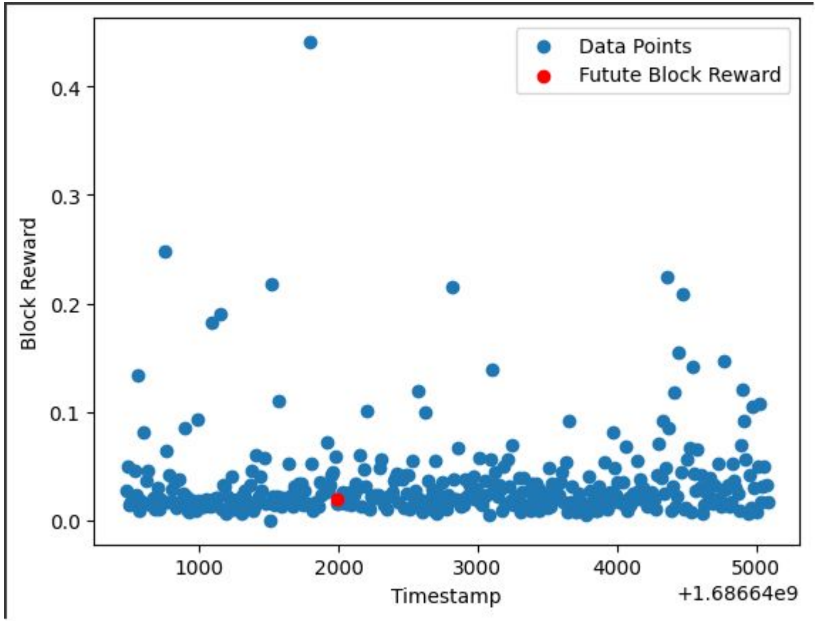
- Initially designed for Ethereum but adaptable to diverse blockchain ecosystems.
- Flexible approach ensures potential application across various blockchain technologies.

7. Ethical Considerations

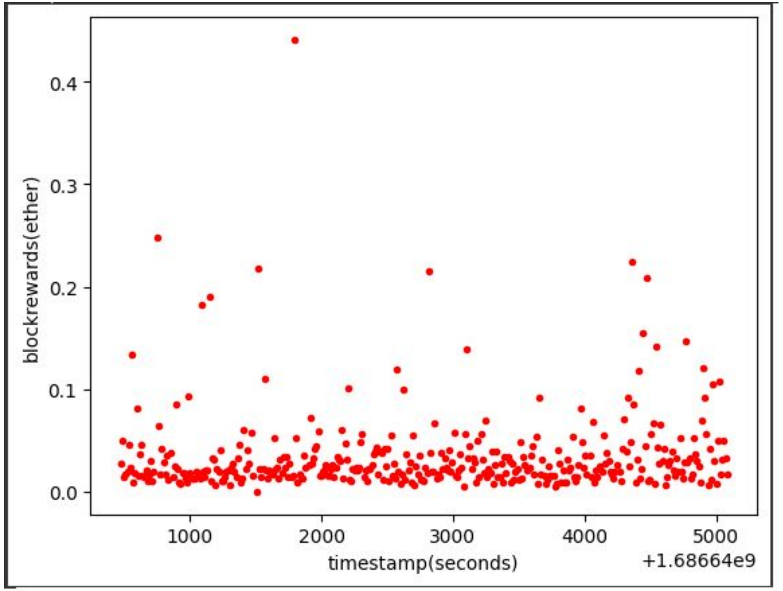
- Prioritized data privacy and confidentiality throughout.
- Deployment adheres to responsible data practices.


Blockchain Platform
API
Real-time Analytics
Data Aanlysis
ML MODEL (Random Forest)
Predicted Output
IOT
Smart devices
Sensors

Results



Data Points
Futute Block Reward
Block Reward
Timestamp
+1.68664e9
0.0
0.1
0.2
0.3
0.4
1000
2000
3000
4000
5000



blockrewards(ether)
timestamp(seconds)
+1.68664e9
0.0
0.1
0.2
0.3
0.4
1000
2000
3000
4000
5000

Result

1. The Random Forest Regressor model effectively predicts blockchain-based block rewards, showcasing robustness and suitability for non-linear relationships.

2. Comparison with Linear Regression, SVR, and Gradient Boosting Regression provides insights into individual strengths and weaknesses for blockchain datasets.

3. Acknowledging limitations, future enhancements may involve incorporating more features and exploring ensemble learning techniques for improved accuracy.

Conclusion

- 1. Our research pioneers the integration of Blockchain and AI to enhance data security, allowing companies to train AI models on vast consumer data without compromising privacy.

- 2. The implementation on Ethereum demonstrates the effectiveness of our solution, which is adaptable to various Blockchain technologies, ensuring decentralized, tamper-resistant data repositories.

- 3. Notably, our Random Forest Regression model showcases remarkable accuracy in predicting block rewards, emphasizing ethical considerations and offering a groundbreaking, privacy-preserving solution for diverse industries' data-driven innovation.

References

[1] Vepakomma, P., Gupta, O., Raskar, R., & Narayanan, A. (2018). Split learning for health: Distributed deep learning without sharing raw patient data. arXiv preprint arXiv:1812.00564.

[2] Guo, C., Rengarajan, V., & Soh, C. (2019). A decentralized blockchain- based data management system. Procedia Computer Science, 156, 265- 274.

[3] Tysowski, P., Le, T. T., & Shorten, R. (2020). Privacy-preserving blockchain-based federated learning. Neural Networks, 128, 297305.

[4] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Kuster, T. (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS) (pp. 1175-1191).

[5] Raj, R., Roy, S., Dey, A., & Adhikari, B. (2019). Blockchain-based privacy-preserving data sharing framework for electronic health records. Journal of Medical Systems, 43(10), 299.

Thank You