Making Smart Contracts Smarter

Syed Badruddoja
Department of Computer Science and Engineering
University of North Texas
Denton, TX, 76207, USA
syedbadruddoja@my.unt.edu

Ram Dantu
Department of Computer Science and Engineering
University of North Texas
Denton, TX, 76207, USA
ram.dantu@unt.edu

Yanyan He
Department of Computer Science and Engineering
University of North Texas
Denton, TX, 76207, USA
yanyan.he@unt.edu

Kritagya Upadhayay
Department of Computer Science and Engineering
University of North Texas
Denton, TX, 76207, USA
kritagyaupadhyay@my.unt.edu

Mark Thompson
Department of Computer Science and Engineering
University of North Texas
Denton, TX, 76207, USA
mark.thompson2@unt.edu

2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC) | 978-1-6654-3578-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICBC51069.2021.9461148

***Abstract*—Blockchain technology develops static smart contracts for decentralized business transactions, lacks dynamic decision-making capabilities that limit the possibilities of ever-increasing demands of modern business applications. Artificial intelligence, a computational prediction platform provides intelligent predictions, actions, and recognition that lacks the ability to hold on to the integrity of the prediction result and requires the help of external authorities to secure the system. Blockchain-based AI prediction can cover the gaps of individual technologies and can mutually benefit from one another to develop a decentralized machine learning architecture that promises to yield better security, automation, and dynamism of the application. This paper proposes a Naive Bayes prediction algorithm to perform prediction with inside blockchain smart contracts that promises to open up more opportunities in the field of Blockchain-AI decentralized applications.**

***Index Terms*—Blockchain, DApp, Smart Contract, Machine Learning, Artificial Intelligence, Naive Bayes**

I. Introduction

Artificial intelligence (AI) is a branch of science that specializes in mimicking decisions that are normally taken by a human being with certain experiences. The intelligence in humans comes from learning history, exploring data, and recognizing patterns [1]. Blockchain is a distributed ledger that contains chained information hashes that are the results of transactions made in a decentralized fashion with consensus protocols between a group of miners [2].

Both AI and blockchain technology have had their success stories in industry and academic research but still many applications today demand much more than what is available. With AI and blockchain each having unique features and advantages, the collaboration between these two technologies could be a boon to both the industrial and academic fields for development, implementation, and operation.

II. Motivation

A. Blockchain Helping AI

Liu et al. [3] mentioned that blockchain can help Machine Learning (ML) in many aspects. Data-model sharing is one of the challenges faced by ML application developers where ownership of data and training models are difficult to control. Blockchain can address these problems with its built-in security feature. Using cryptographic techniques, blockchain can provide data confidentiality, authenticity, and auditability to ensure security and privacy. The immutability feature can ensure that the records are free from any tampering [2].

978-1-6654-3578-9/21/$31.00 ©2021 IEEE

B. AI Helping Blockchain

AI can greatly overcome some inherent requirements and limitations of blockchain that are hindering application development. One of the disadvantages of the distributed ledger of blockchain is that the size of the ledger can become very big if the data input size is very large. ML techniques can preprocess the data with normalization and cleaning before feeding into the blockchain framework so that the overburdening of the ledger is removed. Smart contracts can be made legally viable with the help of ML techniques that has a dynamic nature with the help of natural language processing (NLP) [3].

C. Industrial and Academic Research

DeepBrainChain [4] performs decentralization of compute load to reduce costs. CortexAI [5] created a platform for the open participation of developers to build AI-based blockchain applications. Danku [6] allows anyone to post a dataset and ML model that will be evaluated and rewarded ensuring ownership of models. Several academic works [7]–[10] boost the idea of predictive intelligence with blockchain smart contracts providing, automation, trust, privacy, model sharing, and security enhancements to the mutual collaboration benefits. The current industrial implementations and academic research lack the actual collaborative strategy of performing prediction inside blockchain smart contracts due to the limitation of data types that are available with solidity language [11].

III. Problem Statement

Ethereum [12] blockchain platforms define smart contracts for decentralized applications with a set of rules to be followed before making a transaction available in the ledger. The instructions are mostly static whereas most of the modern applications require dynamism and automation. The automation of ML and blockchain would pave the way for many applications to eliminate any third party to make the predictions more reliable, flawless, and free from any manipulations.

IV. Design Architecture

A. Design Overview

The design involves encoding the Naive Bayes machine learning algorithm inside a smart contract to perform prediction based on a previously trained model. The architecture consists of on-chain and off-chain components where on-chain components involve computation inside the blockchain and off-chain involves computation outside the blockchain. The training of the components is done off-chain and the prediction is performed on-chain. Naive Bayes algorithm can compute the posterior probability of a class given the training data along with prior probability and likelihood probability.

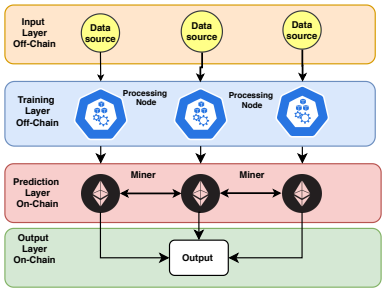

Fig. 1. Layers of ML technique in Blockchain

Figure 2, shows the event flow between different parties in the decentralized application. The off-chain components include the data and training phase for preprocessing the data with feature extraction. Once the training phase is completed, the model can be prepared inside a smart contract with solidity to perform prediction. The on-chain component has two functions to be performed at a high level, probability finder, and class predictor. After completing the process, the prediction is performed and output classification is determined.

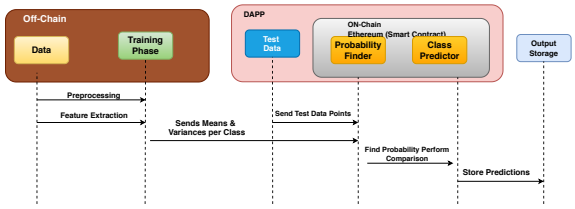

Fig. 2. Event flow in Smart Contract Prediction

B. Naive Bayes Algorithm

The design implementation involves greater work in building the Naive Bayes algorithm for probability calculation inside the smart contract. The computations involved Taylor series expansion [13] derivation that helped to build posterior probability function to compute integer equivalent probability.

V. Implementation & Results

Off-chain: Initially, the dataset was divided into a training and test set. The training set is preprocessed and training weights are recorded that will be used for prediction. In our case, we have taken the means and variances of each feature of each class as a training parameter that will be used by the smart contracts for prediction purposes.

On-chain: The test set of the data is taken along with the means and variances of each class to calculate the probability equivalent value inside the smart contract. We do not find the actual value of probability but perform a comparison of values that can provide an estimate of a probable class out of the two classes compared. In this way, the comparison of all classes is performed in pairs of classes to yield the prediction of the highest probable class.

Dataset: We have considered real datasets for testing our algorithm in a smart contract. The datasets chosen are the iris flowers [14] , pima diabetes [15] and heart disease [17]. The iris dataset originally consists of four features and three classes with 150 samples. The pima diabetes dataset has eight features, two classes, and 750 samples. The heart disease dataset has five features, five classes, and 303 samples.

Result: We have achieved significant accuracy of predictions in smart contract-based implementation when compared with the python based Naive Bayes prediction. As we can see from Figure 3, the iris dataset has an accuracy of 94.66% in the python built-in function and 87.8% in smart contract functions. The pima Diabetes dataset has 78% accuracy in python's built-in function and 67% in smart contract functions. The heart disease dataset has 62% accuracy in python's built-in function while it maintained 58% accuracy in smart contract-based functions.

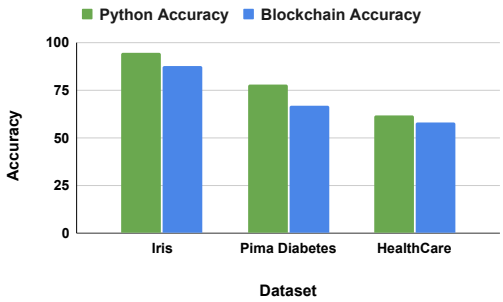

Fig. 3. Prediction Accuracy of real datasets

VI. Conclusion

Though blockchain smart contracts do not support floating point calculation, the numerical expansion with the Taylor series for Naive Bayes algorithm provided promising results to build applications performing the intelligent computation in the blockchain. This opens the possibility to many applications such as DeFi, weather forecast, healthcare, insurance, agriculture, etc. that require the blend of two technologies for highly complex secure decisions to make decentralized predictions

REFERENCES

[1] V. Nigam, “Artificial Intelligence ! What is it actually ?”, https://towardsdatascience.com/artificial-intelligence-what-is-it-actually-7733032083b1, Accessed November 2020.

[2] T. Candido, “A Technical Introduction To Blockchain”, https://medium.com/better-programming/a-technical-introduction-to-blockchain-22ab05308151, Accessed November 2020.

[3] Y. Liu, F. R. Yu, X. Li, H. Ji and V. C. M. Leung, "Blockchain and Machine Learning for Communications and Networking Systems," in IEEE Communications Surveys & Tutorials, vol. 22, no. 2, pp. 1392-1431, Secondquarter 2020, doi: 10.1109/COMST.2020.2975911.

[4] D. Wang, C. Chang, J, Pai,B. Xu, H.Gu,S Liu, K. Ye,Artificial Intelligence Computing Platform Driven By BlockChain, https://www.deepbrainchain.org/assets/pdf/DeepBrainChainWhitepaper_en.pdf

[5] Cortex labs.2018. "AI Smart Contracts — The Past, Present, and Future",December 6, 2018, Retrieved September 3, 2020 from https://medium.com/cortexlabs/ai-smart-contract-5018dc56e2d8

[6] A. B. Kurtulmus and K. Daniel, Trustless machine learning contracts; evaluating and exchanging machine learning models on the Ethereum Blockchain, 2018, [online] Available: https://arXiv:1802.10185. https://github.com/algorithmiaio/danku

[7] J. D. Harris and B. Waggoner, "Decentralized and Collaborative AI on Blockchain," 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA, 2019, pp. 368-375, doi: 10.1109/Blockchain.2019.00057.

[8] Tao Wang and Xinmin Wu and Taiping He.2019."Trustable and Automated Machine Learning Running with Blockchain and Its Applications", KDD 2019 Auto ML

[9] H. Kim. S. Kim. J. Y. Hwang and C. Seo.2019. "Efficient Privacy-Preserving Machine Learning for Blockchain Network" ,in IEEE Access, vol. 7, pp. 136481-136495, 2019, doi: 10.1109/ACCESS.2019.2940052.

[10] Jiameng Liu, Shaoliang Peng, Chengnian Long, Lijun Wei, Yunhao Liu, and Zhihui Tian. 2020. Blockchain for Data Science. In Proceedings of the 2020 The 2nd International Conference on Blockchain Technology (ICBCT’20). Association for Computing Machinery, New York, NY, USA, 24–28. DOI:https://doi-org.libproxy.library.unt.edu/10.1145/3390566.3391681

[11] Solidity v0.7.4- Documentation, https://docs.soliditylang.org/en/v0.7.4/abi-spec.html#types

[12] R. Cordell,”Intro to Ethereum”, https://ethereum.org/en/developers/docs/intro-to-ethereum/#what-is-ethereum, Accessed November 2020

[13] Taylor Series, " https://en.wikipedia.org/wiki/Taylor_series",Retrieved March 10,2021

[14] IRIS dataset, https://archive.ics.uci.edu/ml/datasets/iris

[15] PIMA Indian diabetes dataset https://www.kaggle.com/uciml/pima-indians-diabetes-database

[16] Heart Disease Dataset https://archive.ics.uci.edu/ml/datasets/heart+disease

[17] N. Statt. 2018. “ The AI boom is happening all over the world, and it’s accelerating quickly”,Retrieved from https://www.theverge.com/2018/12/12/18136929/artificial-intelligence-ai-index-report-2018-machine-learning-global-progress-research

[18] Alessandro Mario Lagana Tosch. 2019."Understanding How Artificial Intelligence Can Make Blockchain Safer and Smarter"(February 12,2019, Retrieved September 2020 from https://hackernoon.com/understanding-how-artificial-intelligence-can-make-Ethereum-safer-and-smarter-d8d4c1f4d343

[19] J. Weng. J. Weng. J. Zhang. M. Li. Y. Zhang and W. Luo. 2019. "DeepChain: Auditable and Privacy-Preserving Deep Learning with Blockchain-based Incentive",November 8,2019, in IEEE Transactions on Dependable and Secure Computing, doi: 10.1109/TDSC.2019.2952332

[20] Ricardo Marcacini.2019."How to use machine learning algorithms as Oracles in Smart Contracts?",July 1, 2019, Retrieved September 3,2020 from https://medium.com/artificial-intelligence-for-Ethereum-smart/how-to-use-machine-learning-algorithms-as-oracles-in-smart-contracts-238c6353526a

[21] N. Kumar, J. Madhuri and M. ChanneGowda.2017. "Review on security and privacy concerns in Internet of Things," 2017 International Conference on IoT and Application (ICIOT), Nagapattinam, pp. 1-5, doi: 10.1109/ICIOTA.2017.8073640.

[22] A. Munshi, N. A. Alqarni and N. Abdullah Almalki.2020."DDOS Attack on IOT Devices" .3rd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2020, pp. 1-5, doi: 10.1109/ICCAIS48893.2020.9096818.

[23] S. Dey.2018 "Securing Majority-Attack in Blockchain Using Machine Learning and Algorithmic Game Theory: A Proof of Work," 2018 10th Computer Science and Electronic Engineering (CEEC), Colchester, United Kingdom, pp. 7-10, doi: 10.1109/CEEC.2018.8674185.

[24] Q. Shafi and A. Basit.2019. "DDoS Botnet Prevention using Blockchain in Software Defined Internet of Things," 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), Islamabad, Pakistan, 2019, pp. 624-628, doi: 10.1109/IBCAST.2019.8667147.

[25] K. Salah. M. H. U. Rehman. N. Nizamuddin and A. Al-Fuqaha.2019. "Blockchain for AI: Review and Open Research Challenges" in IEEE Access, vol. 7, pp. 10127-10149, 2019, doi: 10.1109/ACCESS.2018.2890507.

[26] Justin D. Harris.2019. "Leveraging Ethereum to make machine learning models more accessible",July 12, 2019,Retrieved September 3,2020 from https://www.microsoft.com/en-us/research/blog/leveraging-Ethereum-to-make-machine-learning-models-more-accessible/

[27] M. Shen. X. Tang, L. Zhu. X. Du and M. Guizani.2019. "Privacy-Preserving Support Vector Machine Training Over Blockchain-Based Encrypted IoT Data in Smart Cities," in IEEE Internet of Things Journal, vol. 6, no. 5, pp. 7702-7712, Oct. 2019, doi: 10.1109/JIOT.2019.2901840.

[28] T. Wang.2018. "A Unified Analytical Framework for Trustable Machine Learning and Automation Running with Blockchain" 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4974-4983, doi: 10.1109/BigData.2018.8622262.

[29] R. Mauri, “Blockchain for fraud prevention: Industry use cases”, https://www.ibm.com/blogs/blockchain/2017/07/blockchain-for-fraud-prevention-industry-use-cases/, Accessed November 2020

[30] R. Wolfson, “Weather-tracking blockchain in West Africa, but transparency on a raincheck”, https://cointelegraph.com/news/weather-tracking-blockchain-in-west-africa-but-transparency-on-a-raincheck, Accessed November 2020.

[31] Z. Alhadhrami, S. Alghfeli, M. Alghfeli, J. A. Abedlla and K. Shuaib, "Introducing blockchains for healthcare," 2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA), Ras Al Khaimah, 2017, pp. 1-4, doi: 10.1109/ICECTA.2017.8252043.

[32] J.Tsui, “Four Ways Blockchain Is Changing the Food Supply Chain”, https://www.supplychainbrain.com/blogs/1-think-tank/post/30764-four-uses-of-blockchain-in-the-food-supply-chain ,Accessed November 2020

[33] Matrix Technical Whitepaper, Sep. 2018, [online] Available: https://www.matrix.io/html/MATRIXTechnicalWhitePaper.pdf.

[34] W. Rahman, "Scaling AI: 5 Reasons Why It’s Difficult",https://towardsdatascience.com/scaling-ai-5-reasons-why-its-difficult-6ea77b9f7d48

[35] I. Pavlenko, "Blockchain Scalability: Hard Forks, Lighting Network, and Plasma Cash",https://applicature.com/blog/Blockchain-technology/Blockchain-scalability#: :text=Summary,the%20number%20of%20users%20increases

[36] Rohith Gandhi.2018. "Naive Bayes Classifier",May 5,2018,Retrieved September 3,2020, https://towardsdatascience.com/Naive-bayes-classifier-81d512f50a7c

[37] S. Badruddoja, R. Dantu, L. Widick, Z. Zaccagni and K. Upadhyay, "Integrating DOTS With Blockchain Can Secure Massive IoT Sensors," 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), New Orleans, LA, USA, 2020, pp. 937-946, doi: 10.1109/IPDPSW50202.2020.00156.

[38] A. S. Muttavarapu, R. Dantu and M. Thompson, "Distributed Ledger for Spammers’ Resume," 2019 IEEE Conference on Communications and Network Security (CNS), Washington, DC, USA, 2019, pp. 1-9, doi: 10.1109/CNS.2019.8802789.

[39] L. Widick, I. Ranasinghe, R. Dantu and S. Jonnada, "Blockchain Based Authentication and Authorization Framework for Remote Collaboration Systems," 2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Washington, DC, USA, 2019, pp. 1-7, doi: 10.1109/WoWMoM.2019.8792994.

[40] K. Upadhyay, R. Dantu, Z. Zaccagni and S. Badruddoja, "Is Your Legal Contract Ambiguous? Convert to a Smart Legal Contract," 2020 IEEE International Conference on Blockchain (Blockchain), Rhodes, Greece, 2020, pp. 273-280, doi: 10.1109/Blockchain50366.2020.00041.