Nguyen-Hoang, Tu Anh; Hoang, Ngoc Cu; Hua, Phu Thien; Thi, Mong Thy Nguyen; Ta, Thu Thuy; Nguyen, Thu; Tan-Vo, Khoa; Dinh, Ngoc Thanh; Nguyen, Hong Tri

Advancing Scholarship Management : A blockchain-Enhanced Platform with Privacy-Secure Identities and AI-Driven Recommendations

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2024.3486078

Published: 01/01/2024

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Nguyen-Hoang, T. A., Hoang, N. C., Hua, P. T., Thi, M. T. N., Ta, T. T., Nguyen, T., Tan-Vo, K., Dinh, N. T., & Nguyen, H. T. (2024). Advancing Scholarship Management : A blockchain-Enhanced Platform with Privacy-Secure Identities and AI-Driven Recommendations. *IEEE Access*, *12*, 168060-168090. Article 3486078. https://doi.org/10.1109/ACCESS.2024.3486078

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 7 September 2024, accepted 15 October 2024, date of publication 24 October 2024, date of current version 21 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3486078

RESEARCH ARTICLE

Advancing Scholarship Management: A Blockchain-Enhanced Platform With Privacy-Secure Identities and AI-Driven Recommendations

TU-ANH NGUYEN-HOANG[1,2], (Member, IEEE), **NGOC CU HOANG**[1,2], **PHU THIEN HUA**[1,2], **MONG-THY NGUYEN THI**[1,2], **THU-THUY TA**[1,2], **THU NGUYEN**[1,2], **KHOA TAN-VO**[1,2], **NGOC-THANH DINH**[3], AND **HONG-TRI NGUYEN**[4]

[1]Faculty of Information Science and Engineering, University of Information Technology, Ho Chi Minh City 720400, Vietnam
[2]Vietnam National University, Ho Chi Minh City 700000, Vietnam
[3]The Industrial University of Ho Chi Minh City 700000, Vietnam
[4]Department of Computer Science, School of Science, Aalto University, 02150 Espoo, Finland

Corresponding authors: Tu-Anh Nguyen-Hoang (anhnht@uit.edu.vn) and Hong-Tri Nguyen (hong-tri.nguyen@aalto.fi)

This research was supported by The VNUHCM-University of Information Technology's Scientific Research Support Fund.

ABSTRACT Traditional scholarship management systems are often marred by inefficiencies, a lack of transparency, and significant concerns regarding data privacy. These challenges hinder the equitable distribution of educational funds and obscure the scholarship allocation process. This study aims to address these issues by proposing a blockchain-based scholarship management platform. However, scalability issues are introduced along with blockchain-based solutions, for which zk-rollup offers a promising solution. The integration of Self-Sovereign Identity Zero-Knowledge Proof assures secure and private submission of scholarship applications, safeguarding student data while maintaining transparency in the verification process. Additionally, a machine learning model is employed to predict scholarship qualification, functioning as a recommendation system that identifies and prioritizes deserving students. This data-driven approach proactively eliminates barriers that potentially impede deserving students from accessing financial aid, such as administrative oversights or a lack of self-assurance in their qualifications. Our experimental findings confirm the effectiveness of zk-rollups in enhancing transaction efficiency, demonstrating a reduction in transaction confirmation time by approximately 63.6% and a decrease in transaction costs by nearly 90%. Besides, the machine learning model achieved a good performance rating, achieving a balanced accuracy of 86.75% and a mean average precision of 91.68% on a realistically imbalanced test set, reflecting real-world conditions.

INDEX TERMS Scholarship management, blockchain, scalability, zero-knowledge, zk-rollup, machine learning.

I. INTRODUCTION

In modern education systems, scholarship endowment management plays a vital role in maintaining sufficient financial resources for commendable students to continue their education [1]. Scholarships provide educational opportunities for dynamic and outstanding individuals and are a powerful tool to encourage and motivate personal and social development.

Traditional scholarship funding management faces many difficulties and limitations [2]. Conventional methods often involve complex, expensive processes and are prone to identifying and distributing funding errors. Jared Smith, an expert in the education sector, emphasizes the importance of transparency in schools, noting that it fosters trust and

The associate editor coordinating the review of this manuscript and approving it for publication was Berdakh Abibullaev.

engagement, which are crucial for improvement. However, he warns that if not properly managed, transparency can lead to misuse of information by staff.[1] This is particularly concerning in areas like scholarship management, which often involve financial matters. To solve these problems, blockchain technology [3] has appeared and brings outstanding advantages. Blockchain provides a distributed and secure platform that allows data to be stored in linked and unmodifiable blocks. Users can easily track and audit grant distribution, confirming that deserving students receive the financial resources needed to further their education. This means creating an equal learning and development environment for all students, leading to the overall progress of society and education. A successful example of blockchain implementation is seen in the University of Malta's initiative,[2] which tracks and manages scholarship funds, reducing the risk of fraud and attesting to the fact that only qualified students receive scholarships. This case demonstrates how blockchain can enhance trust, improve efficiency, and minimize errors in scholarship management.

However, the application of blockchain technology to scholarship management faces scalability challenges. Although blockchain provides high transparency, confirming transactions on the blockchain network can cause significant delays. This raises the issue of system scalability when a large number of transactions need to be processed simultaneously. This can lead to delays in validation and recording, affecting user experience and system performance. Therefore, the combination with Layer 2 (L2) solution zk-rollup [4] has been proposed, which is especially important to improve the scalability of the system. Zk-rollup allows grouping a large number of transactions into a block and creates a concise proof of the block. Then, smart contracts on Layer 1 (L1) simply verify that proof without re-executing transactions. By using zk-rollup, the system can process a large number of transactions while maintaining transparency.

The Federal Trade Commission[3] and Apply,[4] a platform that helps students find colleges and scholarships, warn students about scholarship scams, particularly those asking for unusual and sensitive personal information, which can lead to privacy breaches and data misuse. Current digital identity management models, which depend on centralized data storage and identity providers, have resulted in an increasing number of data breaches. These breaches have led to significant personal data loss and considerable costs for all stakeholders, especially the users [5]. Inadequate data ownership and control by users over their digital identity data, coupled with the lack of sufficient digital identities for more than a billion people globally [6], significantly undermine privacy rights and hinder access to services [7]. Our proposed smart scholarship management (SSM), offers a solution by implementing a self-sovereign identity (SSI) approach, which empowers students with control over their digital identifiers.

Natalie Keller, a Student & Program Success Specialist, points out that many students hesitate to apply for scholarships due to misconceptions, such as believing scholarships are only for top students or being unaware of available institutional scholarships. These misunderstandings often prevent talented students from pursuing valuable opportunities.[5] Machine learning (ML) can optimize scholarship fund allocation by leveraging its advanced learning capabilities to identify deserving students. Combining blockchain with ML has significant potential to create a secure, decentralized, and efficient administrative system [8], [9]. This study investigates the possibility of integrating blockchain and ML within our innovative platform to predict scholarship qualification effectively. This approach aims to transform the process of awarding and administering scholarships. By merging the transparency, security, and immutability of blockchain with the predictive capabilities of ML, this approach promises to create a robust and efficient system. By analyzing student data and monitoring the progress of scholarship recipients over time, ML algorithms can identify students at risk of falling below scholarship eligibility requirements. SSM enables institutions to implement proactive interventions, such as targeted tutoring or academic counseling, to support these students and foster their continued academic success. Furthermore, ML can provide valuable insights into the correlations between scholarship awards and student retention rates. By leveraging this knowledge, institutions can optimize their scholarship programs, maximizing their impact on student success and promoting long-term educational attainment.

The remainder of this paper is organized as follows: Section II provides the necessary background, while Section III reviews related literature. Section IV details the proposed method, Section V describes the experimental setup and results, and Section VI engages in a discussion about the proposed platform. Finally, Section VII presents the conclusions drawn from this study.

II. PRELIMINARIES

A. SCHOLARSHIP MANAGEMENT

A scholarship is a form of financial aid that does not require repayment, typically awarded to students based on academic achievement, financial need, or other specified criteria [10]. Scholarship administration refers to the processes used to organize and administer scholarships, from the initial notification and application stages to the selection of recipients and distribution of funds. The scholarship management process is described in Figure 1.

[1] drjaredsmith.com/post/transparency-a-double-edged-sword

[2] welcome-center-malta.com/post/e300k-cryptocurrency-scholarship-launched-by-the-university-of-malta

[3] consumer.ftc.gov/articles/how-avoid-scholarship-and-financial-aid-scams

[4] apply.com/guidance/articles/paying-for-college/beware-of-scholarship-scams

[5] linkedin.com/pulse/why-students-arent-applying-scholarships-natalie-keller

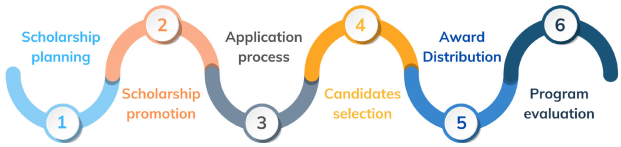

FIGURE 1. **Scholarship management process.**

Scholarships are like a powerful "fertilizer" in the field of education, spectacularly nurturing talent and development. It helps ease the financial burden, allowing promising students to focus on their studies and achieve success. Scholarships act as a bridge between social and economic gaps, facilitating diversity and enriching the academic environment. They give wings to ambitions, motivating students towards their full potential. By helping individuals pursue their academic dreams, scholarships not only shape individual lives but also contribute to a more vibrant and advanced society.

An academic institution will act as a non-profit organization to bridge between donors who want to support students and aspiring students who need financial assistance. Non-profit organizations are businesses that allow individuals to make money but not distribute it to their shareholders [11].

B. BLOCKCHAIN TECHNOLOGY

Blockchain technology, a decentralized transaction and data management technology, first emerged with the Bitcoin cryptocurrency in 2008. Its core characteristic is a tamper-resistant and immutable ledger, distributed across all participants, eliminating the need for central authorities or intermediaries [12]. Blockchain's potential extends far beyond cryptocurrencies, touching various domains like finance, supply chain management, healthcare, and education, due to its attributes that ensure security, transparency, and integrity of data without centralized control [13], [14], [15].

One fundamental aspect of blockchain technology is its basis in hashing, which creates a unique, irreversible output for any given input. This principle underpins the security and trust mechanisms in blockchain systems, enabling the secure recording of transactions and smart contracts on digital platforms without the risk of downtime, censorship, or fraud [16].

The exploration of blockchain applications has led to innovations beyond financial transactions, including securing digital identities, streamlining supply chain operations, and facilitating transparent governance mechanisms. This diversification of use cases highlights the versatility of blockchain and its potential to underpin a wide range of fundraising, economic, social, and organizational activities, fundamentally changing the way how to conduct transactions and manage data in the digital age [17].

C. SMART CONTRACTS

The concept of "smart contract" was first introduced by Nick Szabo in the mid-1990s [18]. He proposed converting contractual clauses into code and embedding them in software or hardware. This would enable contracts to self-execute, thereby reducing the costs of transactions between parties and mitigating the risks of accidental breaches or deliberate interference during the execution of the contract [19].

Smart contracts are automated and self-executing programs that carry out the terms and conditions of a specific contract or agreement using software source code and computational infrastructure [20]. Smart contracts are programs that are stored on a blockchain, similar to other blockchain transactions [21]. It expands distributed ledger technology, operating as decentralized programs within the blockchain network. These programs are immutable and have been cryptographically verified, ensuring their trustworthiness. Important characteristics of smart contracts include direct execution between parties without the intervention of a central third party and the ability to provide services without central dependence. Automation in execution, combined with pre-defined conditions, makes these contracts "smarter" than simply paper-based conditions. The flexibility of smart contracts is inherited from the features of the underlying blockchain technology.

D. ZERO KNOWLEDGE

Zero-knowledge proofs (ZKPs) are a revolutionary concept in the field of cryptography, enabling one party, the prover, to demonstrate to another party, the verifier, that a given statement is true without revealing any information apart from the validity of the statement itself. This concept enhances privacy and plays a crucial role in secure communication and authentication processes across various digital platforms.

The foundational work of Goldwasser, Micali, and Rackoff in the 1980s [22] introduced the formal definitions and conditions under which ZKPs can be considered valid, laying the groundwork for numerous applications in cryptography and computer security. Since then, the development of ZKPs has been closely tied to advances in computational complexity theory and practical encryption technologies, with significant contributions from researchers such as Blum, Feldman, and Micali [23] who developed practical implementations and protocols enhancing the usability of ZKPs in real-world applications.

The security and effectiveness of ZKPs have been the subject of more investigation. For instance, zk-SNARKs (Zero-Knowledge Succinct Non-Interactive Argument of Knowledge), which are notable is the brevity of the proof size and the minimal interaction between the prover and verifier. This advancement significantly enhances the practical applicability of ZKPs in blockchain technologies and privacy-preserving computations [24].

To accommodate different computational settings and assumptions, zero-knowledge protocols have been extended to support various complexity classes and cryptographic assumptions. Groth and Sahai's [25] work on efficient constructions for bilinear groups is particularly notable, providing a framework that is both flexible and powerful, and applicable across a range of cryptographic tasks.

Recent developments have focused on making ZKPs more accessible and scalable. Maller et al. [26] presented a novel construction that reduces the computational overhead and interaction complexity, making ZKPs more feasible for large-scale applications such as secure multiparty computations and private information retrieval systems.

E. ZK-ROLLUP

Zk-Rollups is a technology used in blockchain to make transactions faster and cheaper by grouping multiple transactions and processing them as one. Imagine a group of people pooling their money together to buy a large item at once, instead of each person buying a small item individually. Despite this progress, current blockchain systems still do not achieve the same quality of service as centralized systems, especially in terms of system performance. It requires significant communication resource provision to run effectively [27]. This limitation arises because each network node participating in the Proof-of-Work consistency mechanism must confirm each transaction, requiring significant computing power, storage capacity, and a substantial waiting period. Rollups [4] provide a solution by minimizing the resources and the waiting period required for transaction confirmation. This is achieved using a second-layer network consisting of transaction processing actors outside the main chain. The transaction data is then aggregated into batches and published to the first-layer blockchain. Figure 2 depicts how L1 and L2 interact in the context of a Rollup. Zero-knowledge (ZK) rollups achieve enhanced performance by shifting transaction execution from L1 to L2. Unlike applying the ''innocent until proven guilty'' approach, the ZK rollup aggregators provide proof of authenticity of the original state published with transaction data, ensuring accuracy in calculating transaction execution. This proof of validity is carefully constructed using cryptographic tools such as ZK-SNARK or ZK-STARK [28].

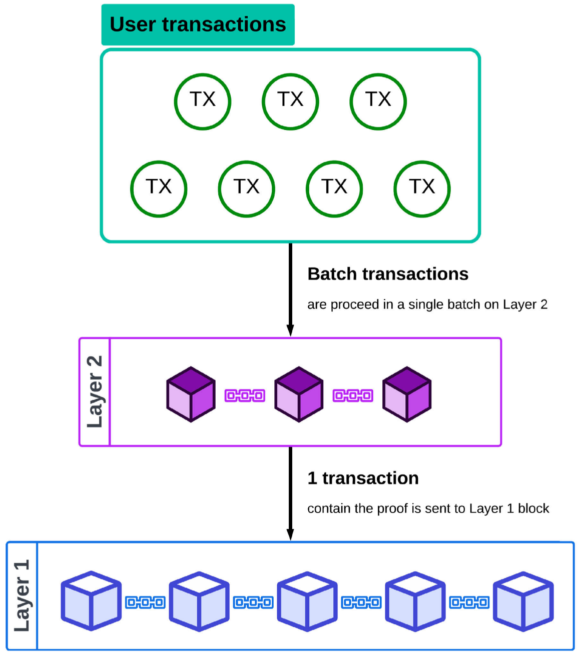

FIGURE 2. ZK-rollup build.

F. SELF-SOVEREIGN IDENTITY

SSI is a digital identity model where individuals have complete control over their personal data and share it directly with others without relying on a central authority. Instead of having multiple ID cards from different organizations, individuals manage their identity and credentials on their own digital platforms. This approach not only improves privacy but also allows individuals to share specific portions of their personal information with service providers as needed, based on the principles of decentralization and user-centricity [29], [30].

SSI systems are built around several core components, including decentralized identifiers (DIDs) [31], verifiable credentials (VCs) [32], and identity wallets. These elements facilitate the creation, storage, and sharing of digital identities without reliance on a central authority. The interaction between these components occurs through a framework where credentials are issued by trusted entities and stored securely on the user's device, allowing for controlled sharing with verifiable authenticity. The Trust Over IP framework is useful for understanding and implementing SSI. It expands on the traditional TCP/IP model to include layers that ensure trust and security in digital interactions. This framework categorizes trust into several layers, ranging from participant identification to identity verification, and emphasizes the importance of trust anchors for establishing secure digital relationships [33].

Recent studies have focused on expanding the SSI model to address various practical challenges, such as the design of robust and interoperable systems for document verification and the need for efficient identity verification processes that can adapt to different application areas. These studies highlight the ongoing development and refinement of SSI technologies to meet evolving digital security and privacy requirements [34].

G. SCHOLARSHIP QUALIFICATION PREDICTION

The process of determining scholarship qualification and allocation has historically been a resource-intensive task that involves evaluating numerous criteria, including academic achievements, financial need, extracurricular activities, and more. The advent of ML and Data Mining (DM) techniques has opened new avenues for automating and enhancing this process, making it more efficient and objective. Recent studies in the field of educational data mining have demonstrated the potential of ML and DM algorithms in various applications, including predicting student performance and dropout rates, and, relevant to this study, scholarship qualification.

ML, a subset of Artificial Intelligence (AI), leverages algorithms to parse data, learn from it, and then make a determination or prediction about something in the world. In the context of education, ML techniques can significantly contribute to identifying patterns and insights within educational data, thereby enabling the development of predictive models for scholarship eligibility. Hegde et al. [35] provided a comprehensive review of the use of ML and DM methods for scholarship prediction, highlighting the effectiveness of algorithms such as Naive Bayes, Decision Tree, and K-NN in accurately determining scholarship candidates.

The predictive analysis of student data not only helps in identifying eligible scholarship recipients but also provides insights that can be used to tailor educational programs and support systems to meet the diverse needs of the student population. As educational institutions continue to seek more data-driven approaches to support student success, the role of ML in scholarship prediction becomes increasingly pertinent.

III. LITERATURE REVIEW

A. FINANCIAL SUPPORT WITH BLOCKCHAIN TECHNOLOGY

The intersection of monetary support and blockchain technology represents a significant advancement toward transparency, accountability, and efficiency in financial aid. This literature review synthesizes findings from various research papers to understand the impact of blockchain technology on financial support systems.

A blockchain-based crowdsourcing loan platform for funding higher education in developing countries [36] introduces an innovative blockchain-based framework to facilitate higher education funding in developing countries. Asamoah et al. propose a decentralized platform where investors can fund students' education through loans. This system not only enables students to access higher education regardless of their financial background but also provides a mechanism for investors to earn returns on their investments. The platform ensures transparency and accountability through the use of blockchain technology, making it a trustworthy and efficient model for educational funding.

Blockchain fundraising and charity platform [37] highlights the critical issue of transparency in traditional charity collection systems. Govindarajan et al. introduced a blockchain-based crowdfunding platform to enhance transparency in the fundraising process and expenses, thus overcoming the trust deficit observed in current systems.

Multi-point fundraising and distribution via blockchain [38] by Alassaf et al. further emphasizes the need for transparency and trust in charities. Their study proposed a system that uses smart contracts to create a transparent fundraising platform, significantly enhancing the integrity and transparency of charity organizations.

Decentralized fundraising application using blockchain [39] addresses the donor's concerns regarding the utilization of funds. Dange et al. developed a decentralized application that provides secure, transparent, and easy use of funds, promoting a more significant amount of donations by enhancing donor trust through blockchain transparency.

Research on charity systems based on blockchain [40] and charity systems based using blockchain [41] both explore the implementation of a charity system on the Ethereum platform. Hu and Li, and Anju and Gini discuss how blockchain technology can increase the transparency of charities, thus fostering public trust and encouraging philanthropy development.

A donation-based system using blockchain [42] by Mhatre et al. proposes a system that addresses the misuse of donated funds through blockchain technology. This system enables direct contributions to specific causes, with transactions recorded on a public ledger, enhancing the accountability and transparency of charitable giving.

A platform for tracking donations of charitable foundations based on blockchain technology [43] presented by Saleh et al. focuses on developing a single platform that provides transparent accounting and monitoring of donations, transactions, and donor information, further bolstering the trust of donors.

A secured distributed ledger-based fundraising framework using smart contracts [44] and understanding users' reaction to blockchain technology on the online fundraising platform [45] provide insights into user perspectives and the operational frameworks of blockchain-based fundraising platforms, underscoring the importance of technical credibility and the design features of blockchain in enhancing user engagement and trust.

Integrating blockchain technology into financial support systems significantly enhances transparency, trust, and efficiency. The reviewed literature demonstrates blockchain's potential to transform charitable giving, addressing long-standing issues such as transparency, accountability, and the direct allocation of funds to intended causes. As blockchain technology continues to evolve, its application in financial aid is poised to foster a more trustworthy, transparent, and efficient financial assistance program.

B. SCHOLARSHIP FUND MANAGEMENT WITH BLOCKCHAIN TECHNOLOGY

Blockchain technology can revolutionize scholarship fund management by enhancing transparency, security, and efficiency through its immutable ledger and smart contracts. It enhances trust among stakeholders and facilitates timely, automated fund distribution.

TEduChain [46], developed by Rashid et al., presents a decentralized platform to crowdsource tertiary education funds, establishing transparent and tamper-proof records with an automated investment process validated by a prototype in Java 8 SE. This platform facilitates contracts between students and sponsors, managed by fundraisers acting as miners in the blockchain network, showcasing the versatility of blockchain applications beyond traditional charity.

ScholarChain [47], proposed by Tekgüç Adalier, and Yurtkan, utilizes smart contracts for transparent and efficient scholarship management within Turkey's educational system, automating verification and payment processes.

Smart contract-based central sector scheme of scholarship for college and university students [48], by Bedi et al., introduces a blockchain-based smart contract system for the Central Sector Scheme of Scholarship for College and University Students in India. This system enhances transparency, efficiency, and security in the scholarship management process by automating verification and payment processes and integrating four key entities (students, education boards, colleges, and banks) within the blockchain framework. The system's robustness was tested against various cyber-attacks, demonstrating its security except for certain vulnerabilities, and proposes future improvements for broader accessibility.

Jadhav et al. present Scholar Block [49], a blockchain-based system designed to improve Corporate Social Responsibility (CSR) through better scholarship management. Using Hyperledger Fabric, the framework creates a transparent, secure, and efficient platform for managing scholarship donations. The system uses two chaincodes: CSR funds for company-to-NGO (where NGO stands for non-governmental organization) transfers and scholarships for NGO-to-student disbursements. CouchDB handles local data storage, while IPFS manages secure documents. The platform addresses transparency, fraud, and inefficiency in philanthropy, aiming to reduce operational costs, streamline resources, and build trust among stakeholders.

CryptoScholarChain [50], by Swati and Pise, offers an Ethereum-based framework addressing transparency and efficiency issues in traditional systems. It integrates smart contracts and a secure payment mechanism, validated through a prototype, to manage scholarship applications, disbursements, and tracking comprehensively.

A secure and scalable approach for scholarship funding management based on blockchain technology with zk-rollups [51], by Hoang et al., presents a blockchain-based solution that leverages zk-rollups to enhance scalability while warranting security and transparency in scholarship management. This implementation shows marked enhancements, reducing transaction costs by up to 25 times and boosting processing speed by about 53%.

These platforms collectively improve transparency, security, and efficiency in managing scholarships, showcasing the transformative potential of blockchain technology in higher education funding. However, they have yet to address scalability issues and the protection of student privacy during the scholarship application process.

C. STUDENTS' QUALIFICATION FOR SCHOLARSHIP PREDICTION

The determination of students' qualifications for scholarships involves assessing a multitude of factors. Recent research leveraging data mining and decision-making methodologies offers new insights into optimizing scholarship allocation processes. This literature review explores several studies that have applied various techniques to predict scholarship qualification, enhancing fairness, and efficiency in scholarship awards.

Application of fuzzy multi-attribute decision-making method [52] by Ilham et al. explores the use of the Fuzzy Multi-Attribute Decision Making (FMADM) method to predict scholarship qualification. This method considers multiple criteria, such as academic performance, parental income, number of siblings, and parents' education. The study demonstrates FMADM's utility in making objective and precise scholarship allocation decisions, providing a framework that can significantly benefit educational institutions in the scholarship distribution process.

Classification models for higher learning scholarship award decisions [53] by Ahmad and Bakar employs data mining techniques to develop a classification model for scholarship award decisions. Utilizing algorithms such as J48, SVM, NB, ANN, and RT, the study evaluates each model's efficiency in predicting scholarship eligibility. The findings suggest that SVM and NB algorithms provide the most accurate classifications for "Eligible" and "Not Eligible" statuses, respectively. This research contributes to the scholarship selection process by offering a data-driven approach to determining scholarship awards.

Decision support system scholarship selection using simple additive weighting (SAW) method [54] by Arifitama examines the application of the SAW method in the scholarship selection process. This study addresses the challenge of subjective decision-making in scholarship awards by providing a systematic and objective framework. The SAW method enables the evaluation of applicants based on predefined criteria, ensuring that scholarships are awarded to the most deserving students based on a transparent and equitable process.

Advancements in predictive analytics and decision-making methodologies offer significant opportunities to improve scholarship allocation processes. By leveraging data mining techniques and decision-support systems, educational institutions can promote fair and efficient distribution of scholarships, aligning with merit and need. The reviewed literature underscores the potential of these technologies to transform scholarship award decisions, making them more data-driven and objective.

D. COMPARE WITH CURRENT WORKS

Table 1 presents a comparative analysis of related platforms on scholarship fund management systems incorporating blockchain technology. This comparison aims to provide a clear understanding of how blockchain technology has been utilized in this field and the various obstacles that need to be addressed for wider implementation. Each study is evaluated based on key strengths and weaknesses, such as transparency, security, privacy, and efficiency.

Table 2 delves into a comparative analysis of existing research on decentralized scholarship management. It summarizes various studies investigating how blockchain technology can revolutionize the field by enhancing transparency, accountability, and efficiency in managing scholarships. However, these studies have yet to address significant obstacles such as scalability and privacy challenges that must be overcome to achieve greater efficiency for successful implementation.

Existing decentralized scholarship management solutions utilize blockchain technology and smart contracts on the EVM to establish a secure, transparent, and decentralized framework. These innovations attest that contributions and disbursements are traceable and immutable, while also improving efficiency by reducing the need for intermediaries, lowering transaction costs, and expediting the contribution process. However, EVM faces practicality challenges due to the EVM's limited performance, as highlighted by Vitalik Buterin [55]. Building upon this foundation, Hoang et al. [51] addressed scalability limitations through the implementation of zk-rollup, achieving a 95.98% transaction fee reduction and a 53.28% faster transaction process compared to the EVM-based approach, laying the groundwork for greater uptake.

Our proposed solution extends this progress further by incorporating several key advancements:

SSM utilizes a zero-knowledge Ethereum Virtual Machine (zkEVM) [56], combining the EVM with ZKPs, similar to the approach used by Hoang et al. [51]. This approach addresses the scalability limitations of the EVM by bundling multiple transactions into a single proof through the zk-rollup technique. Consequently, This approach reduces network congestion, lowers transaction fees by 89.98%, and increases transaction speed by 63.6%, all while maintaining robust security and decentralization. As a result, this approach enables SSM to operate efficiently on a larger scale, facilitating wider adoption.

The zkEVM and SSI ensure smart contract execution and student privacy, safeguarding financial transactions regarding scholarships and user data stored on the blockchain. Additionally, SSI benefits students by streamlining the authentication and eligibility verification processes.

Moreover, our work employs ML to optimize fund allocation by identifying the most deserving students and providing tailored recommendations, while also exploring the integration of blockchain and ML in scholarship fund management.

SSM aims to advance the field of decentralized scholarship management while building upon the valuable contributions of previous research. These enhancements collectively contribute to a wider perspective, scalable, privacy-focused, and efficient scholarship management system. However, the expanded scope of our work introduces potential challenges, such as complex implementation and difficulties in integrating with existing systems.

IV. PROPOSED METHOD

This section presents the structural design for integrating blockchain technology and ML in a scholarship management system.

Figure 3 presents a detailed illustration of the platform's architecture, showcasing the integral components and their interconnections. To enhance performance, cost efficiency, and user experience, it's crucial to discern between on-chain and off-chain operations for execution. On-chain transactions boast immutability, transparency, and security but may incur higher costs and longer processing times characteristic of blockchain technology. These operations are best suited for:

- Scholarship Fund Management: Direct management of scholarship funds on the blockchain guarantees that the allocation of funds is transparent and secure, encompassing the receipt of donations, custody of funds in smart contracts, and their distribution to beneficiaries.
- Scholarship Awards and Claims: Documenting scholarship awards and claims on the blockchain creates a permanent record of awardees, aiding in fraud prevention. This involves smart contracts that disburse funds upon the fulfillment of specific criteria.
- Identity Verification: Students would possess and manage their own identity. Blockchain technology and ZKPs enable individuals to verify their identity without revealing any personal information through SSI solutions.

Blockchain is not a one-size-fits-all solution as its transparency can conflict with the "right to be forgotten" and data rectification requirements. Conversely, off-chain operations, which are more cost-efficient and quicker, are preferable for tasks demanding extensive computation, large-scale data storage, or enhanced privacy. These operations are ideally employed for:

- Application Processing: The process of managing scholarship applications, including the collection of personal details and documents, is best-conducted off-chain to protect data privacy and handle data volume effectively.
- Evaluation and Review: Reviewing applications and making scholarship decisions require significant computational efforts. Performing these tasks off-chain can be more cost-efficient.

The on-chain layer offers a transparent, immutable, and secure framework for managing scholarship funds, contributions, and distributions. Meanwhile, the off-chain layer deals with the processing of scholarship applications and employs a machine-learning model to forecast scholarship qualifications, offering recommendations for students. To seamlessly integrate these two layers, a robust data bridge

TABLE 1. Strengths and weaknesses of related platforms.

Related solution	Strengths	Weaknesses
[46] (2020)	TEduChain presents a novel approach to education funding by creating a decentralized platform that connects students, fundraisers, and sponsors. Blockchain secures transparency and immutability of transactions, increasing trust in the crowd-funding process. The system's ability to automatically generate and manage contracts between students and sponsors is an advantage over traditional systems.	There is limited discussion on how the system would ensure the privacy of sensitive student information while maintaining transparency. Additionally, blockchain scalability challenges may potentially arise.
[47] (2020)	This work proposes using blockchain and smart contracts to create a transparent and automated scholarship management platform. Key strengths include increased transparency in the scholarship granting process, automated payments through smart contracts, and a tamper-proof record of all transactions.	This work does not discuss potential scalability issues that could arise when implementing this system for a large number of scholarships and students. Additionally, it does not address privacy concerns that may come with storing sensitive student information on a blockchain. The proposed solution appears to be in its conceptual stages.
[48] (2020)	This research proposes a blockchain-based smart contract system to manage India's Central Sector Scheme of Scholarship. Its strengths include improved traceability of applications, elimination of physical document transit issues, enhanced transparency between students and education boards, and automated verification of bank accounts. The system aims to reduce manual labor and time consumption in the scholarship process.	This work lacks a discussion on the scalability and privacy concerns associated with implementing this system for a vast number of scholarships and students.
[49] (2023)	This blockchain-based system offers transparent and secure scholarship fund management through smart contracts and a decentralized ledger. It integrates corporate social responsibility initiatives to expand funding sources and uses Hyperledger Fabric for improved control and efficiency in a permissioned environment.	This work does not delve into the potential scalability issues or privacy risks that could arise when applying this system to a large-scale scholarship program.
[50] (2023)	This framework leverages blockchain and smart contracts to enhance transparency, traceability, and efficiency in scholarship management. It integrates multiple funding sources, including corporate social responsibility initiatives and government programs. IPFS helps address data integrity and accessibility challenges in document storage.	The framework does not yet address the scalability challenges associated with the Ethereum Virtual Machine (EVM), which could result in high transaction costs during network congestion. Additionally, the framework does not fully explore the potential privacy implications of storing sensitive student data on a public blockchain.
[51] (2023)	This research introduces an approach combining blockchain with zk-rollups, addressing both transparency and scalability concerns. Compared to EVM-based systems, the L2 scaling solution offers transaction processing that is 53.28% quicker and reduces transaction costs by 95.98%. The inclusion of a governance mechanism allowing donors to participate in fund allocation decisions is an innovative feature that enhances stakeholder engagement.	This work does not fully explore the potential trade-offs that could arise from relying on the L2 and privacy issues.

is in place, ensuring secure and efficient data exchange. This integration guarantees that information regarding students eligible for specific scholarships from the off-chain layer is accurately transmitted to the on-chain layer, facilitating efficient disbursement of funds.

A. SYSTEM COMPONENTS

1) STUDENTS

When a scholarship fund becomes available for applications, students have the opportunity to apply for financial aid. If students fulfill all the requirements of a particular scholarship fund, they qualify for that scholarship. A student on the platform is represented as Tuple 1:

$$S = \left\{ dId, sId, loc, natId, aiId, walletAddress, t \right\} \tag{1}$$

where:

- *dId*: The public DID of the student.
- *sId*: The student's unique identifier assigned by their academic institution.
- *loc*: The location of the student.
- *natId*: The student's national identifier.
- *aiId*: The identifier of the student's enrolled academic institution.
- *walletAddress*: The wallet address of the student.
- *t*: The time of registration of the student.

2) ACADEMIC INSTITUTIONS

Academic institutions lead the charge in scholarship campaigns, planning and launching them on the blockchain to establish a transparent, immutable foundation for funding efforts. Moreover, academic institutions serve as a bridge between donors and students, engaging with donors to

TABLE 2. Comparison between SSM and other related platforms.

Related solution	Virtual Machines	Decentralization	Security	Privacy	Scalability	ML/AI
[46] (2020)	EVM	✓	✓	✗	✗	✗
[47] (2020)	EVM	✓	✓	✗	✗	✗
[48] (2020)	EVM	✓	✓	✗	✗	✗
[49] (2023)	EVM	✓	✓	✗	✗	✗
[50] (2023)	EVM	✓	✓	✗	✗	✗
[51] (2023)	zkEVM	✓	✓	✗	✓	✗
SSM - our work	zkEVM	✓	✓	✓	✓	✓

✓: Addressed
✗: Not addressed
EVM: Ethereum Virtual Machine
zkEVM: zero-knowledge Ethereum Virtual Machine

highlight the impact of their contributions and guiding students through the application process. This is because the institution possesses the authority to manage and uphold smart contracts, playing a critical role in ensuring the seamless operation and integrity of the scholarship ecosystem.

3) DONORS

Donors serve as the financial backbone that supports the educational aspirations of students. Through their contributions, donors enable the realization of scholarship campaigns, directly facilitating access to education for deserving students. They can observe, in real time, the progress of funding campaigns toward their targets, reinforcing the tangible impact of their donations. Moreover, blockchain technology allows donors to witness the fulfillment of campaigns, from the achievement of funding targets to the eventual disbursement of scholarships to students' blockchain wallet addresses. This end-to-end visibility into the scholarship lifecycle not only ensures donors that their contributions are serving their intended purpose but also showcases the efficiency and security of using blockchain technology in philanthropy.

4) SCHOLARSHIP CAMPAIGNS

Academic institutions are set to launch a scholarship campaign, which will be financially supported through donor cryptocurrency contributions. To guarantee transparency, every aspect of the campaign's creation and the distribution of funds will be documented on a blockchain. The campaign will collect funds and provide financial aid to deserving students upon achieving its funding goals.

Each scholarship campaign is represented as Tuple 2:

$$C = \Big\{ cId, aiId, title, description, target, amountCollected, \\ donors, donations, deadline, status, sCriteria, createdAt \Big\} \tag{2}$$

where:

- *cid*: The identifier of the campaign.
- *aiId*: The identifier for the academic institution requesting to launch the campaign on the platform.
- *title*: The official name of the scholarship campaign.
- *description*: A brief overview of the campaign.
- *target*: The goal amount of cryptocurrency to be raised.
- *amountCollected*: The current sum of cryptocurrency contributions received by the campaign.
- *donors*: An array that records the wallet addresses of each donor, uniquely identifying the source of every contribution to the campaign.
- *donations*: An array that documents the amount of each donation, detailing the specific contribution size from each donor.
- *deadline*: The current sum of cryptocurrency contributions received by the campaign.
- *status*: Indicates the current phase of the campaign.
- *sCriteria*: A detailed set of scholarship criteria established by academic institutions that students must meet to be considered for scholarships once the campaign funding target is achieved and the application period starts.
- *createdAt*: The date and time when the campaign was initially launched

5) SSM - THE SCHOLARSHIP MANAGEMENT PLATFORM

SSM is a dynamic platform that serves as a central hub, connecting key stakeholders within the ecosystem. It facilitates donors in making contributions to scholarship campaigns initiated by academic institutions. These campaigns are designed to gather financial support for students who merit scholarships, seamlessly bridging the gap between benefactors and beneficiaries.

6) BLOCKCHAIN

a: LAYER 2 NETWORK

L2 refers to a network that operates on top of an L1 blockchain to improve its scalability and efficiency. L2

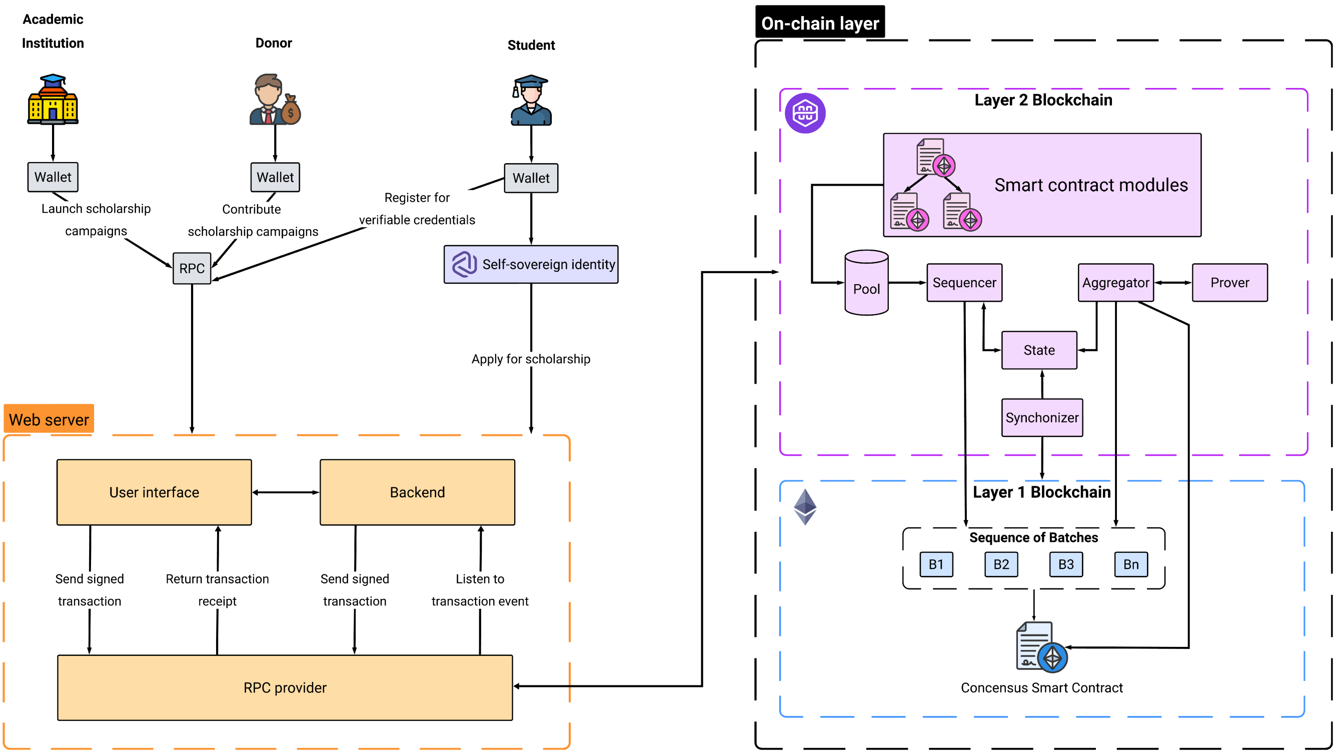

FIGURE 3. Architecture of SSM - the blockchain-based scholarship management.

solutions are protocols that do not require changes to the base layer itself but provide additional infrastructure to handle transactions off the main chain.

Most of the smart contracts that manage the SSM's business logic are deployed on L2. These L2-deployed smart contract modules, as outlined in Figure 3, include:

- Student Identity Contract: This smart contract manages the crucial link between a student's unique identifier and their blockchain wallet address. It facilitates student authentication within the SSI ecosystem and establishes a secure, transparent channel to disburse scholarship funds.
- Academic Institution Registration Contract: This contract maintains a registry of trusted academic institutions. It implements a rigorous verification process, allowing only authenticated institutions to initiate scholarship campaigns. This protects the legitimacy of scholarship offerings and maintains the platform's integrity.
- Scholarship Fund Management Contract: This sophisticated contract oversees the entire lifecycle of scholarship campaigns, including fund allocation, disbursement, and tracking. It implements transparent fund movement protocols, providing real-time visibility into scholarship financials for all stakeholders.
- Application Processing Contract: This contract enhances the efficiency of the scholarship application process. While the actual processing of applications occurs off-chain, the contract records the content identifier from IPFS associated with each student's ID and wallet address. It provides secure storage and management of student applications, implementing data encryption and access control mechanisms to protect sensitive information while facilitating efficient review processes.

Users on the L2 blockchain interact with the platform through Remote Procedure Call (RPC) services. Their transactions are subsequently recorded in the Pool Database (Pool DB). A designated node, known as the sequencer, is responsible for retrieving transactions from the Pool DB. The sequencer then performs validation checks to verify transaction legitimacy before assembling them into optimized batches. These validated batches are subsequently submitted to the L1 network by the sequencer. Finally, the sequencer sequences the submitted batches, assuring their eventual inclusion within the L1 blockchain state.

Complementing the sequencer is the aggregator, another critical node. The aggregator's primary function is to generate cryptographic proofs, specifically ZKPs, verifying the integrity of the state changes proposed by the sequencer. To achieve this, the aggregator utilizes a sophisticated cryptographic tool called the prover. The prover possesses the remarkable capability of simultaneously generating ZKPs for numerous batches. Even more impressive, it can then combine these individual proofs into a single, concise ZKP. This is made possible by Plonky2,[6]

[6]github.com/0xPolygonZero/plonky2/blob/main/plonky2/plonky2.pdf

a cryptographic library that includes a SNARK implementation utilizing techniques from FRI (Fast Reed-Solomon Interactive Oracle Proofs) and PLONK (Permutations over Lagrange-bases for Oecumenical Noninteractive arguments of Knowledge), with a focus on fast recursive methods. The strategy for achieving this succinctness is illustrated in Figure 4.

The process begins with the Executor, which interprets EVM bytecodes using a new zero-knowledge assembly language to establish polynomial constraints that each valid batch of transactions must satisfy, generating the committed polynomials. Once the Executor converts the transactions and associated data into committed polynomials, the STARK Recursion component initiates the production of recursive zk-STARK proofs. This component generates multiple zk-STARK proofs, gathers them into bundles of zk-STARK proofs, creates a further zk-STARK proof for each bundle, and ultimately produces a single final zk-STARK proof. This proof is then processed by the Circom Library, which constructs an arithmetic circuit (expressed as Rank-1 Constraints System) and generates a witness. Finally, the zk-SNARK Prover takes the witness and the STARK verifier data (instructions to process data and generate a succinct zk-SNARK proof) to produce a succinct zk-SNARK proof. This consolidated proof, embodying the validity of all the batched transactions, is then published to the L1 Consensus Contract for verification.

This approach leverages the speed and trustless setup of zk-STARKs alongside the succinctness of zk-SNARKs, creating an efficient and cost-effective zero-knowledge proving system that appreciably reduces gas costs for state change validations. According to Polygon Zero Team,[7] this approach can reduce gas costs from 5 million to 350,000.

b: LAYER 1 NETWORK

L1 refers to the base layer of a blockchain network. This is the foundational layer where the blockchain itself operates. It includes the core protocols defining block creation and validation, the consensus mechanism (such as Proof-of-Work or Proof-of-Stake), and the native cryptocurrency transactions.

The L1 blockchain hosts a crucial smart contract known as the Consensus Contract. Its primary function is to validate the validity proofs generated by the aggregator. Through this verification process, the Consensus Contract guarantees the integrity of all state transitions within the L2 network, ensuring their faithful execution. The Consensus Contract only verifies the proof without having to check each individual transaction. This process dramatically reduces the computational load on the L1 blockchain and allows the network to process many more transactions than it could natively.

[7] docs.polygon.technology/zkEVM/architecture/zkprover

B. SELF-SOVEREIGN IDENTITY IN SSM

Academic institutions and students are identified using unique DIDs [31], a form of globally recognized, persistent identification. These cryptographically generated and registered DIDs are anchored on a decentralized ledger, such as a blockchain, thus guaranteeing secure and immutable references to each entity's digital identity. This method eliminates the need for centralized validation, safeguarding privacy and security in identity management.

All sensitive information related to students is conveyed through VCs, which provide a secure way for students to share their information for authentication on the platform. Physically, credentials include identification details of the holder (e.g., photo, name, or ID number), information about the issuing authority (e.g., a municipal, national, or certifying organization), and so forth. Hence, credentials are assertions made by an issuer. A VC enhances this by being resistant to tampering and featuring cryptographically verifiable authorship, enabling the formation of verifiable presentations with similar security features. While encapsulating the same information as physical credentials, integrating technologies like digital signatures elevates the integrity and credibility of VCs over their physical counterparts.

$$VC = \Big\{ aiId, aiName, sId, dId, walletAddress, documentation, validThru, expiresEnd \Big\} \tag{3}$$

In the domain of SSI and its application to decentralized scholarship management, Tuple 3 formalizes the structure of a VC that students have to present when applying for scholarships. This VC will be utilized to generate a ZKP using the Groth16 protocol, a zk-SNARK scheme known for its small proof sizes and constant-time verification. The proof acts as cryptographically secure, tamper-evident digital attestations, verifying a student's academic eligibility and other pertinent information. By enabling the selective disclosure of verifiable information without revealing underlying details, ZKPs enhance privacy and security in the scholarship application process while streamlining and fortifying the entire procedure. The structural components of this VC are as follows:

- *aiId:* A unique identifier for the academic institution (Academic Institution ID) issuing the VC, confirming the authenticity and traceability of the VC to its source.
- *aiName:* The official name of the issuing academic institution. This attribute offers essential context and additional validation of the issuer's identity.
- *sId:* A unique identifier assigned to the student by their academic institution.
- *dId:* the DID associated with the student's VC.
- *walletAddress:* A distinct cryptographic address that enables the secure receipt of cryptocurrency payments directly into the student's digital wallet.
- *documentation* A reference to a set of documents containing the student's relevant academic information.

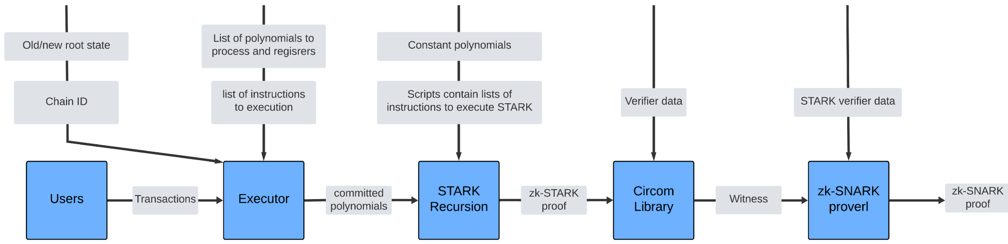

FIGURE 4. **The strategy to attain proof succinctness.**

This includes semester GPA, extracurricular points, academic awards, and other details required for scholarship eligibility assessment.

- *validThru (Valid Through):* The date until which the VC remains valid.
- *expiresEnd (Expiration Date):* The date after which the VC is no longer considered valid.

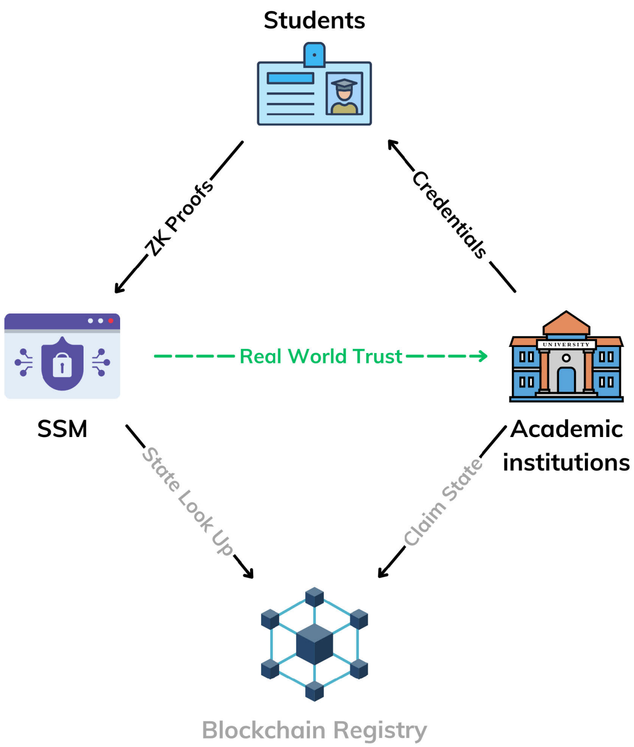

FIGURE 5. **SSI solution with zero-knowledge proofs concept.**

Academic institution issues VCs to their students. Every VC comes from an academic institution and they are cryptographically signed by the corresponding academic institution. Students subsequently store these VCs securely within their digital identity wallets. Figure 5 illustrates the architecture of the SSI solution, which includes three essential components: Students (identity holders), academic institutions (issuers), and SSM - the scholarship management platform (the verifier).

- *Academic institutions*: Academic institution issues VCs to their students. Every VC comes from an academic institution and they are cryptographically signed by the corresponding academic institution.
- *Students*: Upon receiving VC, students cryptographically link them to their DID. To apply for scholarships, students generate ZKPs from their VCs and present these proofs to SSM. SSM verifies the authenticity of the proofs and assesses their eligibility against scholarship criteria. This proof presentation, known as a Verifiable Presentation, allows students to selectively disclose only the necessary information from their VCs, maintaining the confidentiality of other personal data. This selective disclosure process empowers students with greater control over their information privacy.
- *SSM*: SSM verifies the ZKPs presented by the students. In the real world, students would need to provide an ID and show all their personal information. With this SSI solution, they only need to pass a ZKP.

To implement SSI, two key components are required:

- Global Identity State Contract: At the core of the platform's identity management, this smart contract maintains the Global Identity State Tree. This data structure represents a snapshot of all identity states within the system. It enables secure, privacy-preserving authentication of identities through identity profiles, balancing transparency with data protection.
- Trusted Setup: Since the protocol utilizes Groth16, a trusted setup between students and the SSM is necessary. This setup is vital to prevent potential manipulation by SSM and to maintain overall security and trustworthiness. The setup involves:
  - Multi-party computation: Multiple parties contribute randomness (toxic waste) to construct the common reference string (CRS) iteratively used for creating and verifying ZKPs.
  - Key destruction: The randomness used to generate the CRS must be destroyed immediately to prevent SSM from generating false proofs.

Within the sphere of scholarship management, SSI empowers academic institutions to issue VCs regarding their students, students need to provide certain documents to their academic institution for the issuance of a VC. Concurrently, the SSM platform verifies ZKPs generated from these credentials to authenticate student identities during their scholarship application process. This approach leverages ZKPs, allowing students to verify their credential ownership while keeping the underlying details confidential. To initiate the scholarship application process, students first must acquire a VC from their educational institution. With this VC, students can prove their eligibility and apply for scholarships on the platform.

C. MACHINE LEARNING-DRIVEN PRE-ASSESSMENT FOR SCHOLARSHIP QUALIFICATION

ML offers numerous advantages in the context of scholarship management. Still, this study will concentrate on predictive analytics since guaranteeing that qualified students are not unintentionally excluded from the scholarships they deserve is a critical objective in scholarship management. This requires proactive efforts to raise awareness about available scholarships and to debunk misconceptions about student eligibility and competitiveness. ML models offer a promising solution to this challenge. ML can identify potential candidates who meet scholarship criteria by analyzing patterns in historical data on scholarship recipients and their qualifications. This supports the awarding of scholarships to the most praiseworthy students.

Integrating blockchain technology with ML offers a revolutionary method for managing scholarships to enhance security, transparency, and efficiency. Figure 6 further examines this integration. The integration of blockchain and ML in this context is mutually reinforcing. Blockchain guarantees that the data fed into ML models is tamper-proof and authentic, which is crucial for maintaining the integrity of the predictions and recommendations made by the ML algorithms. Blockchain's immutable ledger provides a transparent record of all transactions and data changes, building trust among stakeholders, including students, academic institutions, and scholarship providers. The decentralized nature of blockchain warrants that data is not stored in a single location, reducing the risk of data breaches and ensuring continuity even if part of the network goes down. Smart contracts on the blockchain can automate the execution of scholarship disbursements based on predefined criteria validated by ML models, streamlining the process from application to disbursement, reducing delays, and minimizing manual intervention.

The data sources are foundational to the system, comprising various databases that store crucial information. These data sources are critical as they provide the raw data needed for ML processes:

- **Students:** Contains student profiles, including personal information and academic records.
- **Grades:** Stores detailed grade records, essential for assessing scholarship eligibility and qualifications.
- **Application data:** Holds data related to scholarship applications submitted by students.
- **Student documentation:** Includes necessary documentation such as identification and academic certificates.
- **Scholarships:** Contains details of available scholarships, including eligibility criteria and application deadlines.

SSM employs a strategic hybrid storage approach to mitigate censorship risks while maintaining scalability. Critical, immutable attributes like tamper-proof metadata (e.g., semester academic performance, extracurricular points) are securely stored on the blockchain, leveraging its decentralized nature. Meanwhile, larger files like student documentation for SSI are managed on decentralized storage (IPFS), offering decentralized storage with content addressing for integrity and verification. Centralized storage is strategically utilized to handle the massive volume of detailed student grade records generated each semester, prioritizing efficient access and scalability. This combined strategy optimizes security and performance, customizing the storage solution to the specific needs of each data type within the scholarship management ecosystem.

Feature engineering workflow begins with extracting relevant data from diverse sources, including traditional databases, decentralized storage, and the blockchain. Following extraction, data transformation is performed to clean and structure the data, addressing issues like missing values, outliers, and categorical variables. This process is important for optimizing the data's compatibility and effectiveness with ML algorithms for analysis and prediction. Feature engineering itself involves creating new features or modifying existing ones to enhance the predictive power of the ML models. This includes generating features that effectively represent students' historical academic performance and their engagement in extracurricular and co-curricular activities score. Data ingestion then loads the processed and engineered features (versioned features) into a system where they can be accessed for ML training and predictions.

The ML workflow is the driving force behind identifying meritorious scholarship recipients in this architecture. It begins with data extraction, where the cleaned and engineered features are retrieved for ML tasks Data preparation follows, further refining the data to ensure it is suitable for model training. Data is then split into training, validation, and testing sets, and scaled or transformed as needed. The prepared training data is carefully balanced to achieve fair representation across different scholarship categories and then fed into selected ML algorithms, which learn the underlying patterns and relationships that characterize worthy students. Through iterative refinement, the model's performance is enhanced by fine-tuning hyperparameters and validating its accuracy against unseen data. The validated model is then deployed to predict which new applicants are most likely to be deserving based on the learned patterns. To maintain its effectiveness over time, the model's performance is continuously monitored, and it is retrained with updated data if any degradation is detected.

The proposed architecture offers a progressive approach to scholarship management, creating a fair, transparent, and efficient platform for all stakeholders. By harnessing

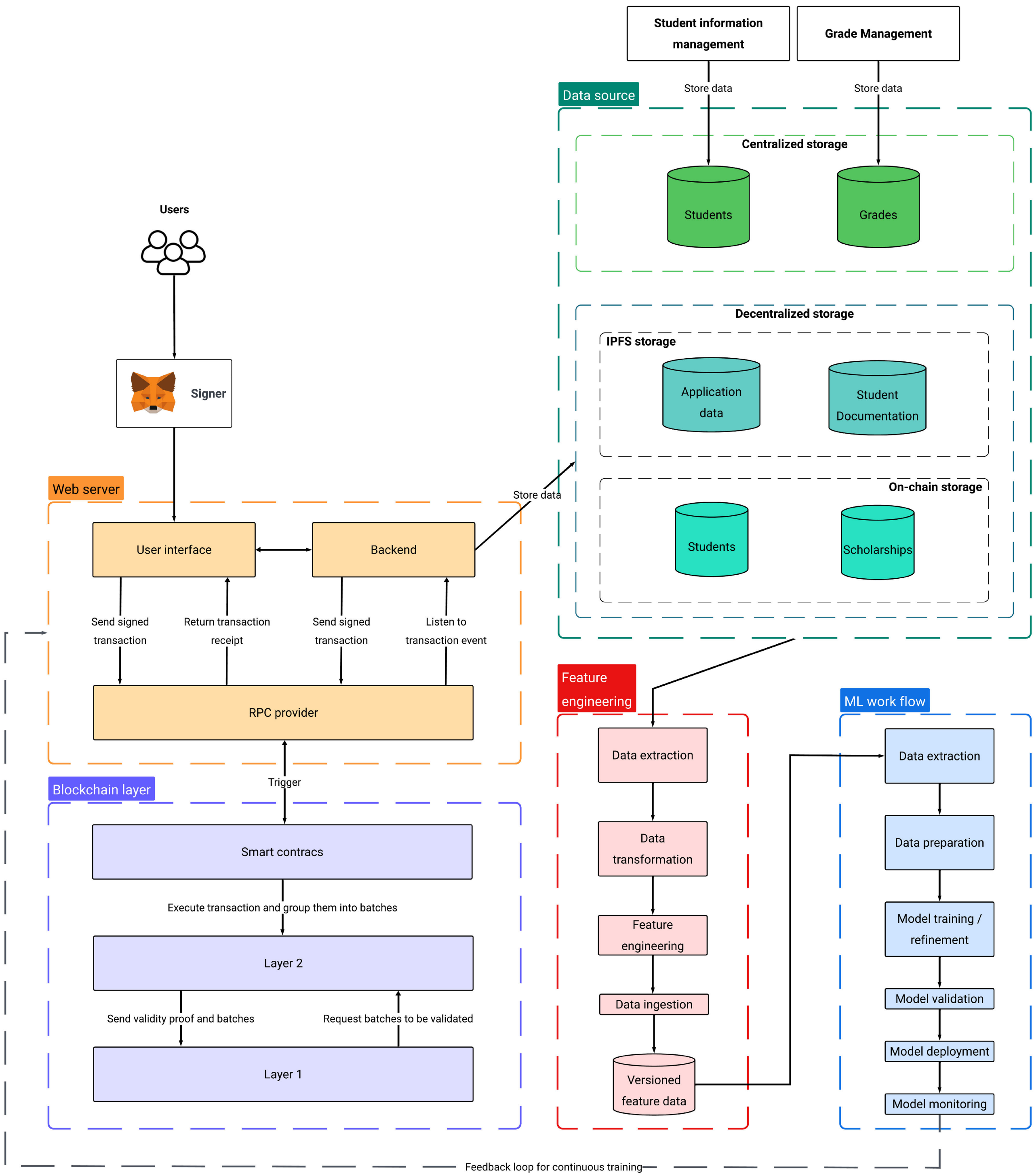

FIGURE 6. **Blockchain and ML integration in scholarship management.**

the immutability and security of blockchain, the system guarantees data integrity and transparency throughout the entire scholarship lifecycle. Meanwhile, the intelligent automation and predictive analytics capabilities of ML streamline administrative tasks and provide valuable insights for proactive interventions, as well as for identifying worthy scholarship recipients.

D. COMMUNICATION DESIGN

Figure 7 serves as a high-level visual representation of the SSM's functionality and its interactions with stakeholders. The platform facilitates interactions among academic institutions, students, donors, and system administrators.

- Academic Institutions initiate the process by launching scholarship campaigns and setting eligibility criteria.

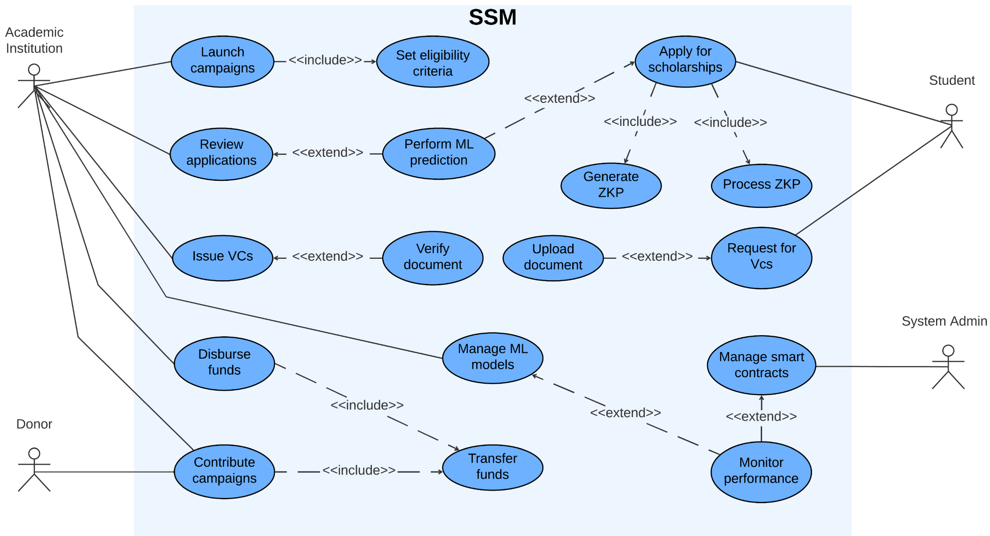

FIGURE 7. Platform's functionality and actor's interactions.

Furthermore, they are responsible for issuing VCs to students and reviewing applications for fund disbursement.

- Students interact with the system by applying for scholarships through SSI, allowing them to request VCs and generate ZKPs to protect their privacy. They also benefit from ML recommendations to discover scholarship opportunities.
- Donors contribute to scholarship campaigns, with smart contracts attesting that their contributions are securely managed and distributed via blockchain.
- System Administrators manage the SSM platform's infrastructure. Their responsibilities include maintaining and upgrading smart contracts and refining ML models for scholarship matching.

Figure 8 depicts the design of application communication, outlined in the subsequent ordered list:

1) *Initiation of scholarship campaigns by academic institutions*: Academic institutions launch scholarship campaigns on the blockchain, establishing specific targets and deadlines. This process is transparent and sets the foundation for funding efforts.
2) *Donor contributions*: Donors contribute to these campaigns by sending funds to a designated blockchain address. The blockchain records each contribution, guaranteeing transparency and the permanence of records.
3) *Achievement of funding targets*: The system constantly monitors the campaign's funding level. When the funding target is met before the deadline, the campaign is declared fully funded, opening the door for student applications.
4) *Verification of student ZKPs*: Before application, students must secure VCs as evidence of their eligibility, such as proof of current enrollment. These credentials are issued by recognized authorities (their academic institution) and encrypted on the blockchain. Rather than submitting VCs directly with their applications, students generate ZKPs from the encrypted credential data. These proofs enable students to demonstrate the authenticity of their credentials and their eligibility for the scholarships they are applying for, while selectively disclosing only the necessary information. This method thereby protects student privacy, enables selective disclosure of necessary information, and minimizes data exposure for ML model training. The model training process will focus exclusively on relevant features that influence scholarship allocation decisions, avoiding the collection of irrelevant, sensitive, highly personal student information.
5) *Scholarship application process*: Students generate a ZKP from their VCs to privately verify their identity and academic qualifications for scholarship applications. This method safeguards sensitive personal information that remains confidential while still allowing students to demonstrate their eligibility.
6) *Application evaluation and selection*: The platform securely records student applications, facilitating evaluation by the review committee. Successful applications are subsequently recorded on the blockchain, ensuring a transparent and tamper-proof record of scholarship awards.
7) *Disbursement of scholarships*: Approved scholarships are directly disbursed to the students' blockchain wallet addresses. This provides a secure, transparent, and tamper-proof transfer of funds, marking the final step in the scholarship awarding process.

E. LOGIC DESIGN

This section offers a closer examination of the platform's key operators.

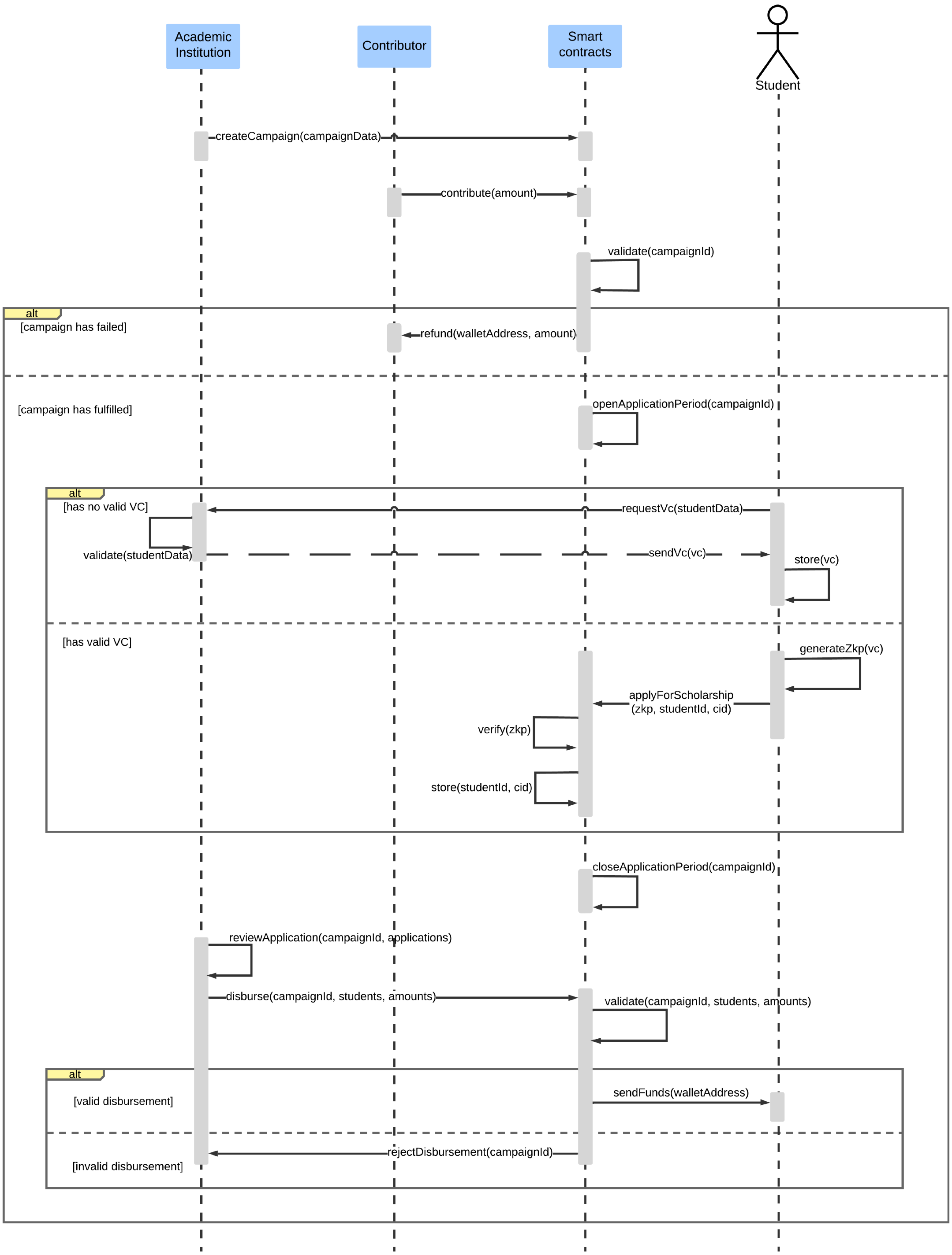

FIGURE 8. The communication flow between system components.

1) ISSUANCE OF VERIFIABLE CREDENTIALS

Algorithm 1 describes the process of issuing VCs from an academic institution. This critical procedure enables students to receive digitally signed credentials that can be independently verified. The process begins with the student's request for a VC, which includes several key inputs:

- *dId*: The student's DID.
- *sId*: The unique identifier for the student requesting the credential.
- *walletAddress*: The digital wallet address where the fund will be sent.
- *documentation*: The set of documents provided by the student to verify their identity and credentials.
- *valdThru*: The start date from which the credential is valid.

- *expiresEnd*: The end date after which the credential expires.

The first step involves verifying the provided documentation. The function verifyDoc(sId, documentation) is called to check the validity of the documentation against the student information. The result of this verification is stored in the variable isValid. If the documentation is deemed invalid (isValid is false), the process terminates early, returning the message "Invalid documentation." This allows only students with valid and verified documentation to proceed further.

If the documentation is valid, the academic institution proceeds to issue the VC by calling the function issueVc. This function creates a verifiable credential using the provided details and the validated documentation, with the result stored in the variable VC. Finally, the VC is returned, allowing students to save the VC to their digital identity wallet.

SSM verifies proofs against the current state of information published on the blockchain. Importantly, it keeps personal data off-chain, only storing the hashed identity state on the blockchain. This hashed identity state is achieved using Merkle Trees.

Algorithm 1 Issuance of Verifiable Credentials

Input: {sId, walletAddress, documentation, validThru, expiresEnd}
Output: {vc}
isValid ← verifyDoc(sId, documentation)
if *isValid* is *false* **then**
  return "Invalid documentation"
end if
vc ← issueVc(dId, sId, walletAddress, documentation, validThru, expiresEnd)
return *vc*

2) SCHOLARSHIP CAMPAIGN MANAGEMENT

Algorithm 2 delineates the procedure for orchestrating scholarship campaigns within a blockchain-based scholarship management platform, enabling efficient and transparent management. The process begins by capturing essential details such as the academic institution identifier (*aiId*), title (*title*), description (*descr*), target funding amount (*target*), deadline for the scholarship campaign (*deadline*), eligibility criteria (*sCriteria*), and timestamp when the campaign is created (*createdAt*). These details are used to create a new campaign through the function *createCampaign*, and the created campaign object is stored in the variable *c*.

The platform continuously monitors the campaign's status, specifically focusing on the campaign's deadline. When the campaign's deadline is reached ($c.deadline \geq now$), the algorithm checks if the campaign has met its financial target. If the campaign is successful, meaning it has met the target funding amount, the system transitions into the application phase by calling the function openForApplication($c.id$), allowing eligible students to apply for scholarships from the successfully funded campaign. Conversely, if the campaign fails to meet the target funding amount, the function refund($c.id$) is called, initiating the prompt return of contributed funds to the donors. This safeguards the donors' contributions and provides a secure refund in the event of an unsuccessful campaign.

This streamlined process leverages blockchain technology to automate the scholarship lifecycle while guaranteeing transparency and immutability of all campaign data and transactions. By using smart contracts and automated triggers, the system maintains the integrity of the scholarship campaign management, providing a reliable and efficient platform for both donors and applicants.

Algorithm 2 Scholarship Campaign Management

Input: {aiId, title, descr, target, deadline, sCriteria, createdAt}
Output:
c ← createCampaign(aiId, title, descr, target, deadline, sCriteria, createdAt)
`When the scholarship campaign ends`
if $c.deadline \geq now$ **then**
  if *c* is *successful* **then**
    openForApplication(c.id)
    setStatus("Open for application")
  else if *c* is *failed* **then**
    refund(c.id)
    setStatus("Refunded")
  end if
end if

3) SCHOLARSHIP CAMPAIGN CONTRIBUTION

Algorithm 3 provides a secure and efficient process for contributing to scholarship campaigns on SSM. Initially, donors must specify the target scholarship campaign (*cId*) and the contribution amount (*amount*). The system automatically captures the donor's wallet address (*donorWalletAddress*). Subsequently, the platform verifies that the contribution meets the minimum required amount, referred to as *MINIMUM_AMOUNT*.

Suppose the contribution amount is equal to or greater than the minimum required. In that case, the system proceeds to securely record the contribution on the blockchain using the function fund(*cId*, *donorWalletAddress*, *amount*). This function assures that the donation is attributed to the specified campaign and securely logs the transaction. Upon successful contribution recording, a receipt (*rcpt*) is generated and returned to the donor, providing proof of the transaction and enhancing transparency.

However, if the contribution amount is less than the minimum allowed, the system immediately provides feedback to the donor by reverting the transaction and displaying the message "Less than minimum amount." This prompt feedback mechanism prevents any insufficient contributions from being processed, thereby maintaining the integrity and efficiency of the donation process.

This process enables the secure and transparent handling of all contributions to scholarship campaigns, leveraging blockchain technology to provide an immutable record of transactions.

Algorithm 3 Scholarship Campaign Contribution

Input: {donorWalletAddress, cId, amount}
Output: {rcpt}
  if $amount \geq MINIMUM\_AMOUNT$ **then**
    $rcpt \leftarrow$ fund(cId, donorWalletAddress, amount)
    return *rcpt*
  else if $amount < MINIMUM\_AMOUNT$ **then**
    revert("Less than minimum amount")
  end if

4) APPLICATION SUBMISSION

To participate in scholarship applications, students must first acquire VCs from the institutions. Students initiate the application process by specifying their DID (dId), student ID (sId), wallet address ($walletAddress$), and the desired scholarship campaign ID (cId). They then submit a VC (vc) relevant to the scholarship requirements to generate a proof (ZKP) for verification. This proof allows students to share only the necessary academic information without revealing their entire record. The platform's automated verification system confirms the authenticity and validity of the VC. Upon successful verification, the platform securely records the student's application details, including their student ID, wallet address, relevant academic information, and the targeted scholarship campaign. Finally, the system provides a confirmation message indicating a successful application submission.

Algorithm 4 Application Submission

Input: {dId, sId, walletAddress, cId, vc}
Output: {applied}
  $zkp \leftarrow$ generateZkp(vc)
  $isValid \leftarrow$ validate(dId, sId, walletAddress, zkp)
  if *isValid* is *false* **then**
    revert("Invalid credentials")
  end if
  $applied \leftarrow$ record(dId, sId, walletAddress, academicInfo, cId)
  return *applied*

5) FUNDS DISBURSEMENT

Algorithm 5 outlines the secure and transparent logic governing scholarship fund disbursement on our blockchain-based platform. Authorized personnel initiates scholarship fund disbursement by specifying the campaign ID (cId), available funds, awarded students (*students*), and individual award amounts (*amounts*). An automated check (*validate*) validates the total disbursement does not exceed the remaining funds. The system securely transfers designated scholarship amounts directly to each student's wallet via blockchain technology if valid. Each transfer is permanently recorded on the blockchain for transparency, and the campaign status is updated to "Paid Out" Finally, the system confirms successful disbursement.

Algorithm 5 Funds Disbursement

Input: {cId, students, amounts}
Output:
  $isValid \leftarrow validate(cId, amounts)$
  if *isValid* is *false* **then**
    revert("Insufficient balance")
  end if
  for $i \leftarrow 0$ to $student.length$ **do**
    transfer(students[i].walletAddress, amounts[i])
  end for
  updateCampaignStatus(cId, "PaidOut")
  return "Funds have been disbursed"

V. EXPERIMENTS

In this section, we will explore the implementation of a blockchain-based scholarship management system and conduct experiments on the student scholarship qualification prediction models.

A. THE BLOCKCHAIN-BASED SCHOLARSHIP MANAGEMENT APPLICATION'S PERFORMANCE ANALYSIS

This section showcases a simulation of the blockchain-based scholarship management platform. It aims to demonstrate user interactions and communication during transaction processing, considering factors like processing time and cost metrics outlined in Table 3. The simulation is conducted on both the Ethereum Sepolia and Polygon zkEVM testnet networks to evaluate the efficiency enhancements provided by L2 scaling solutions in comparison to Ethereum L1.

1) EXPERIMENTAL SETTINGS

In the applied platform, smart contracts are developed using the Hardhat (development environment), written in the Solidity programming language, with a Solidity compiler (solc) version of 0.8.9 or higher. These contracts are deployed on the Sepolia and Polygon zkEVM Cardona networks. Interaction with the smart contracts is facilitated by TypeScript and the Ethers.js libraries, with ThirdWeb[8] serving as the RPC provider. The deployment and interaction processes are carried out on a MacBook Pro, which is powered by the Apple M1 chip, featuring an 8-core CPU and 16GB of RAM.

[8]thirdweb.com

TABLE 3. Platform performance metrics.

Metrics	Fomular	Interpretation
Transaction processing time	$T = t_i + t_c$	Transaction processing time [57] refers to the duration it takes for a transaction to be verified and confirmed on a blockchain network.
Transaction cost	$C = G \times P$	Transaction cost [58] refers to the expense incurred in processing a transaction on a blockchain network.

t_i: Transaction issuance time
t_c: Transaction confirmation time
G: Gas used
P: Gas price

2) EVALUATION METHOD

Key features of the platform include campaign creation, donor contributions, and student fund disbursements. Recognizing the anticipated high volume of contributions, the platform prioritizes the donation process as the primary metric for evaluating transaction efficiency. This choice is driven by the streamlined nature of simulating donations, coupled with minimal constraints, facilitating an accurate assessment of transaction confirmation times and associated costs.

In a controlled environment, a series of transactions were executed across two blockchain networks to assess the efficacy of the L2 scaling solution on L1. The experiment leveraged Sepolia as a simulated L1 network, while Polygon zkEVM Cardona served as the representative L2 solution. This experiment was repeated two times on each network, with every cycle processing a different volume of transactions. Specifically, individual sessions managed groups of 100 and 500 transactions, referred to as epoch 1 and epoch 2, respectively, documenting the time and expense required for the processing of each group.

3) RESULTS

TABLE 4. Comparison between Ethereum Sepolia and Polygon zkEVM Cardona in performance.

		Sepolia	Polygon zkEVM Cardona
Time (hh:mm:ss.mmm)	Epoch 1	00:25:0.340	00:09:6.072
	Epoch 2	01:52:39.195	00:45:36.563
Cost (Gwei)	Epoch 1	8,858,113.939	887,270.786
	Epoch 2	45,243,642.040	4,889,194.673
Cost standard deviation		17,844.989	3,691.719

As evidenced by the data in Table 4, the L2 scaling solution demonstrably improves transaction processing efficiency. This enhancement manifests in both accelerated transaction processing times and reduced transaction fees. Furthermore, this positive trend persists across diverse transaction volumes, thus substantiating the inherent scalability benefits of L2 solutions. Moreover, L2 transaction costs demonstrate superior stability in comparison to L1. This is quantified by a roughly 4.83 times lower cost standard deviation on L2.

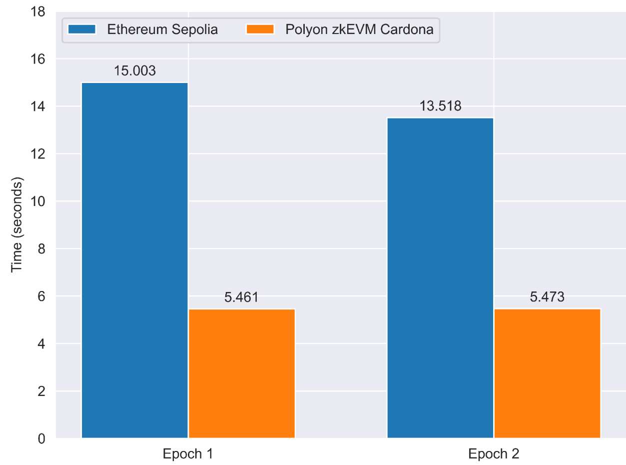

FIGURE 9. Average time for a transaction to be fully confirmed on-chain Comparison between Ethereum Sepolia and Polygon zkEVM Cardona.

Figure 9 visually depicts the transaction confirmation times observed on both the L1 (Sepolia) and L2 (Polygon zkEVM Cardona) networks. As illustrated, the Sepolia network exhibits higher processing times across both epochs. Conversely, the Polygon zkEVM Cardona network displays considerably faster processing times in both epochs. Even with a rise in transaction volume, the corresponding increase in processing time remains comparatively modest, thereby evidencing superior scalability. When compared to L1, the L2 solution achieves a remarkable reduction in transaction confirmation times – by 63.6% in Epoch 1 and 59.51% in Epoch 2.

Figure 10 underscores the cost advantages of the L2 solution. As the data reveals, the Sepolia network incurs higher transaction fees in both epochs. Furthermore, a clear upward trend emerges in Epoch 2, suggesting a linear or even potentially super-linear cost increase with growing transaction volume. In stark contrast, the Polygon zkEVM Cardona network exhibits demonstrably lower transaction costs compared to Sepolia. While costs do increase from Epoch 1 to Epoch 2, they remain remarkably lower than their L1 counterparts, highlighting the cost-effectiveness of the L2 solution even under increased load. Notably, the L2 solution achieves a remarkable reduction in

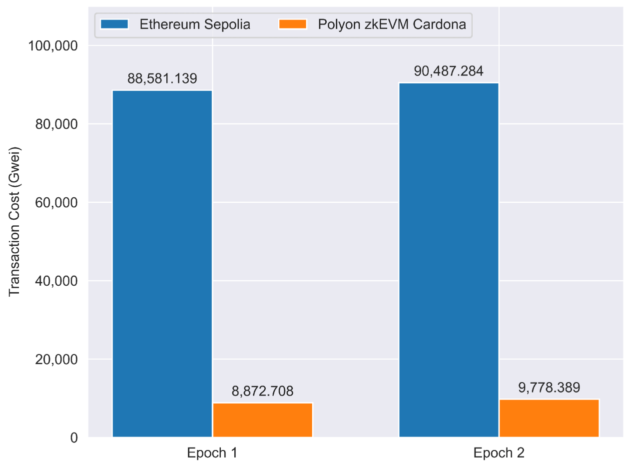

FIGURE 10. Average transaction cost comparison between Ethereum Sepolia and Polygon zkEVM Cardona.

transaction costs – by 89.98% in Epoch 1 and 89.19% in Epoch 2.

This improvement archived because zk-rollups optimize blockchain transactions through a multi-faceted approach. They batch hundreds or thousands of transactions together, reducing per-transaction costs on the main chain. By performing most computations off-chain, they reduce the main chain's burden, enabling parallel processing. Advanced cryptographic techniques compress transaction data, minimizing on-chain storage and processing needs. The use of zero-knowledge proofs allows efficient verification of multiple transactions with a single proof, drastically reducing main chain interactions and associated gas costs. This batching process not only lowers costs but also enables quick finality once the proof is verified, speeding up transaction confirmations. Moreover, zk-rollups exhibit excellent scalability characteristics – as batch sizes increase, per-transaction costs decrease, allowing for efficient scaling as network usage grows. This comprehensive strategy results in significantly reduced transaction costs and increased throughput compared to traditional on-chain processing.

This improvement is achieved because Polygon zkEVM Cardona utilizes zk-rollups to optimize blockchain transactions through a multi-faceted approach. They reduce per-transaction costs on the main chain by batching hundreds or thousands of transactions together. Most computations are performed off-chain, lightening the load on the main chain and enabling parallel processing. Recursive zk-STARK proofs enable the efficient validation of multiple transactions with a single proof, significantly reducing main chain interactions and lowering gas costs. Advanced compression algorithms are applied during the recursive proof generation process to minimize the number of required polynomials, resulting in a more compact proof size. At the end of the proof generation stage, a zk-SNARK proof is selected to replace the zk-STARK proof, aiming to further reduce verification complexity and create a footprint-sized proof. The batching mechanism not only lowers costs but also accelerates transaction confirmations, providing quick finality once the proof is verified. Additionally, zk-rollups scale efficiently—as batch sizes grow, per-transaction costs decrease, making them ideal for handling increased network usage.

The improvement underscores the importance of L2 solutions in enhancing the scalability of blockchains. As the transaction volume increases, the advantages of using L2 solutions like Polygon zkEVM Cardona become even more apparent, making them indispensable for high-throughput blockchain applications.

B. STUDENT SCHOLARSHIP QUALIFICATION PREDICTION

The prepared data and the developed ML model are utilized to predict students' qualifications for the academic encouragement scholarship. This scholarship is awarded regularly to students based on their performance each semester.

1) EXPERIMENTAL SETTINGS

This study examines the Academic Encouragement Scholarship (AES), a merit-based award designed to motivate and inspire top-performing students to excel in their academic pursuits. The criteria for this scholarship usually depend on the student's academic results and personal extracurricular activities over a specific semester.

Students who do not fall into any of the following categories are eligible to apply for the AES:

- Students who have a course with a score below 5.0 in the semester.
- Students who exceed the program completion timeframe (including when students apply for suspension or deferral).
- Students with no tuition fees for the current semester.
- Students who do not participate in Health Insurance.
- Students who are disciplined from the level of reprimand and above in the semester.
- Students who take less than the typical number of credits in the semester.

Following the application closure, all student submissions are categorized by class. A specific number of scholarship slots are then allocated to each class. Scholarships are awarded to the top enrolled students in each class, considering both GPA and extracurricular involvement, until the designated slots for each class are filled. This prioritizes well-rounded students within each class, not the overall applicant pool. These awarded scholarships are subsequently classified into three tiers: Excellent, Very Good, and Good. The specific criteria for each tier, based on semester GPA and extracurricular points, are outlined in Table 5.

In this study, ML is employed to predict student success in acquiring AES at Vietnam National University Ho Chi Minh City (VNUHCM) - University of Information Technology[9] (UIT).

[9]ctsv.uit.edu.vn/sites/default/files/202204/254_qd_dhcntt_22_4_2022.pdf

TABLE 5. AES award levels.

Categories	GPA	Extracurricular point (EP)	Value
Excellent scholarship	GPA $\geq$ 9	EP $\geq$ 90	3
Very good scholarship	GPA $\geq$ 9	90 < EP $\geq$ 80	2
	GPA $\geq$ 8	EP $\geq$ 80	
Good scholarship	GPA $\geq$ 9	80 < EP $\geq$ 70	1
	GPA $\geq$ 8		
	GPA $\geq$ 7	EP $\geq$ 70	
No scholarship	Not listed in the categories above or all scholarship slots have been filled		0

In the preprocessing phase, we began by collecting student performance data from UIT's Student Affairs Office spanning from 2013 to 2022. This data, sourced from both public datasets and internal records, was originally in Excel format. An exploratory data analysis was conducted using Python libraries including Pandas, scikit-learn, seaborn, and matplotlib. This process revealed several data quality issues, which we subsequently resolved.

There exist compulsory courses, but policies allow students to be exempted if they possess valid certifications that meet the course's learning outcomes as defined in the curriculum. This results in missing grade data for these courses as students do not participate due to the exemption. We addressed missing values by assigning the student's GPA from the semester they submitted the certification to these exempted courses. The data we have collected is sourced from a real academic institution and has undergone rigorous validation. Consequently, occurrences of values not conforming to standard formats or duplicate entries are virtually nonexistent.

Subsequently, feature engineering was conducted to extract meaningful information from the data, adhering to UIT AES regulations. Categorical variables were transformed using label encoding, while numerical features were standardized for uniformity. To prepare the dataset for modeling, it was partitioned into training and testing sets using a stratified 70/30 split to preserve class distribution. It is natural that instances where no scholarship was awarded significantly outnumber those where scholarships were granted. To address this class imbalance in the training set, the Synthetic Minority Oversampling Technique (SMOTE) [59] was employed. Finally, Bayesian optimization with a Gaussian process was utilized to optimize hyperparameters for the ML model. This optimization involved fitting a Gaussian process to the objective function and using an exploration strategy, such as Upper Confidence Bound or Expected Improvement, to identify optimal parameter values.

After the preprocessing phase, the dataset offers detailed insights into student achievements and their scholarship qualification within a scholarship management system. This dataset is designed for a classification model aimed at predicting scholarship qualifications and categorizing scholarships among students. This rich dataset, boasting over 78,000 entries, tracks student academic performance, extracurricular involvement, and class information across semesters. This data provides a robust foundation for training and evaluating ML models capable of predicting scholarship qualifications and effectively categorizing suitable scholarships for each student.

Table 6 presents the dataset's details, highlighting its key attributes. For in-depth analysis or model training, especially when evaluating longitudinal student performance or changes in scholarship qualifications over time, it is advisable to index the data by **Student ID**, **Semester**, and **School Year**. This structuring will facilitate a segmented analysis, allowing for a nuanced understanding of each student's academic trajectory and the impact of various factors on scholarship qualification across different periods.

In the model training process aimed at predicting student scholarship qualifications, a subset of features from the dataset, identified as critical for influencing scholarship outcomes, will be utilized. The scholarship status of the student serves as the target variable for the prediction model. A detailed description of the features and the target variable involved in this step is provided below:

- Features:
  - Semester GPA: This numerical feature represents the Grade Point Average (GPA) of a student for a particular semester, scaled from 0 to 10. The GPA is a crucial indicator of academic performance, with higher values reflecting better academic achievements. Since scholarship qualification often hinges on academic excellence, this feature plays a pivotal role in the model.
  - Extracurricular Points: Capped at a maximum of 100, this metric quantifies a student's level of participation and achievement in extracurricular activities. Engagement in extracurricular activities is a highly regarded factor in scholarship decisions, as it provides valuable insight into a student's well-rounded development and commitment beyond academic pursuits.
  - Total Credits: This feature counts the total number of academic credits a student has taken in a particular semester. It is an important factor as it signifies the academic workload and capacity of a student. Certain scholarships may have criteria regarding the minimum number of credits to be taken to maintain qualifications.
  - Whether to have failed a course in the semester (has_failed_course): A binary indicator (True/False) reflecting whether a student has failed at least one course during the semester under review. Failing courses is typically a disqualifier for scholarship qualifications, making this a critical feature for identifying students who meet the minimum academic standards.
  - Student year: Numerical data indicating how many years the student has been enrolled, starting from 1. Reflects the student's progression through their academic career, influencing scholarship

TABLE 6. Dataset's key attributes.

Column	Non-null count	Type	Interpretation
student_id	78,039	int64	A unique identifier for each student, facilitating the tracking of individual academic performance over time.
semester	78,039	int64	These temporal markers distinguish the specific periods of academic assessment, enabling a longitudinal analysis of student progress. Each school year consists of two semesters, thereby offering two opportunities for scholarship awards.
school_year	78,039	int64	
gpa	78,039	float64	Measured on a scale of 0 to 10, this critical metric assesses academic performance, with a GPA below 7 disqualifying students from scholarship qualifications.
extracurricular_point	78,039	int64	This metric effectively gauges a student's extracurricular engagement, which plays a pivotal role in determining their qualifications for scholarships.
total_credits	78,039	int64	Reflects the academic workload managed by the student, indicative of their course engagement level within a semester.
class	78,039	object	Categorizes students by major and enrollment year, serving as a basis for allocating scholarship slots based on class size.
has_failed_course	78,039	bool	A boolean attribute signals whether a student has failed any course during a specific semester, which affects their qualification for a scholarship in that semester.
student_year	78,039	int64	Indicates how many years the student has been enrolled. Students surpassing the standard completion time for university studies will not be qualified for scholarships.
scholarship	78,039	int64	Denotes scholarship qualification and type, with values from 0 to 3 based on GPA and extracurricular points, subject to the condition that qualified students must not have any failed courses and must meet minimum GPA and extracurricular requirements.

qualifications, especially as students beyond their standard completion time are generally not qualified for scholarships.

- Semester: The semester attribute, with values 1 or 2, demarcates the academic year into two halves for students, aligning with their student year to mark specific periods in their educational timeline. This attribute synchronizes course offerings for students sharing a major, fostering a unified academic experience and facilitating peer collaboration. Furthermore, when combined with information about the major and student's year, it offers insight into the anticipated workload for scholarship recipients in that major and specific year of study.
- Class: Categorical data representing the student's class, grouping students by major and enrollment year. This grouping directly influences scholarship allocation, as the number of scholarships for each class is determined based on the corresponding class size. Additionally, this feature could capture patterns within specific classes or majors that influence scholarship distributions. Certain majors or academic years experience disproportionately higher award rates due to factors such as program rigor, enrollment numbers, or donor preferences.

- Scholarship: As previously noted in Table 5, the target variable is categorical and ranges from 0 to 3. A value of 0 signifies that a student is not qualified for a scholarship, while values from 1 to 3 correspond to various scholarship types. These scholarships are awarded based on a combination of factors including GPA, extracurricular points, and other criteria. The classification into different scholarship types allows for a nuanced understanding of scholarship qualifications, reflecting various levels of academic and extracurricular achievement.

To address the inherent challenge of imbalanced class distribution in this classification dataset, the SMOTE is employed to balance the dataset effectively during data preprocessing. Following this, the dataset is divided into training and test sets, with 30% of the data reserved specifically for testing. This division is performed in a stratified manner to preserve the overall distribution of classes within each subset, thereby maintaining the integrity and validity of the model evaluation

During the model training phase, the selected features will be used to train a predictive model. This model will learn to identify patterns and relationships between these features and the scholarship outcomes.

2) EVALUATION METRICS

In this classification problem, the one-against-all (or one-versus-rest) approach [60] is employed to decompose the problem into a series of binary classification tasks.

In binary classification problems, as depicted in the confusion matrix presented in Table 7, data instances are typically expected to belong to either positive or negative categories. Each binary prediction can be classified into one of four categories:

- True positive (TP): The correctly predicted positive outcome.
- True negative (TN): The correctly predicted negative outcome.
- False positive (FP) The negative instance is predicted to be positive.
- False negative (FN): The positive instance is predicted to be negative.

Table 7 provides the basis for deriving various commonly utilized metrics, as illustrated in Table 8, which assess the model's performance with different emphases on evaluation perspectives.

TABLE 7. Confusion matrix for binary classification.

	Actual positive class	Actual negative class
Predicted positive class	True positive (TP)	False negative (FN)
Predicted negative class	False positive (FP)	True negative (TN)

Classification metrics from scikit-learn are utilized for performance assessment. In the case of multi-class classification, we adopt averaging methods for precision, recall, and F1 score calculation, resulting in a set of different average scores in the classification report as referred to in Table 9.

3) RESULTS

Multiple models were constructed to determine the most effective one for this multi-class classification problem. The evaluation of these models is showcased in Table 11 and Figure 11, which compare them across key metrics on a test set comprising 23,412 instances. One standout model is the Extreme Gradient Boosting (XGBoost) classifier, an optimized version of gradient boosting that effectively addresses overfitting, split finding, and missing value handling during the training phase through several efficient techniques [65]. Given its superior performance, we have selected XGBoost as the primary model for implementing our solution. The XGBoost model configuration, determined through Bayesian optimization, is presented in Table 10.

Our analysis reveals XGBoost as the optimal model for predicting scholarship qualifications. This assumption is supported by its remarkable performance across all evaluated metrics:

a: EFFECTIVE MULTI-CLASS CLASSIFICATION

The good Macro F1 score of 0.8634 signifies XGBoost's proficiency in handling multi-class classification tasks like scholarship categories. This contributes to fair and accurate classification of students into each scholarship level.

b: MINIMIZED PREDICTION ERRORS

The outstanding Micro F1 score of 0.9666 translates to a low error rate in predicting the appropriate scholarship level for students. This maintains high precision (finding truly deserving students) and recall (not missing admirable students).

c: BALANCED APPROACH

In a scenario with imbalanced scholarship categories, the Weighted F1 score becomes crucial. XGBoost's impressive score of 0.9664 indicates its effectiveness in handling both majority and minority classes. This promotes a balanced and fair approach across all scholarship categories.

d: HIGH PRECISION WITH ACCURACY

The great mAP score of 0.9168 highlights XGBoost's precision in identifying the correct scholarship level. This translates to maintaining accuracy even when considering a wider range of potential classifications, minimizing the risk of misclassification, and reinforcing the integrity of the scholarship awarding process.

e: GOOD PERFORMANCE ACROSS CLASSES

While not the absolute highest, the Balanced Accuracy score of 0.8675 is still commendable. It indicates that XGBoost performs well in classifying students into both minority and majority scholarship categories. This is vital in scholarship management to avoid any systematic bias towards or against any particular group of students.

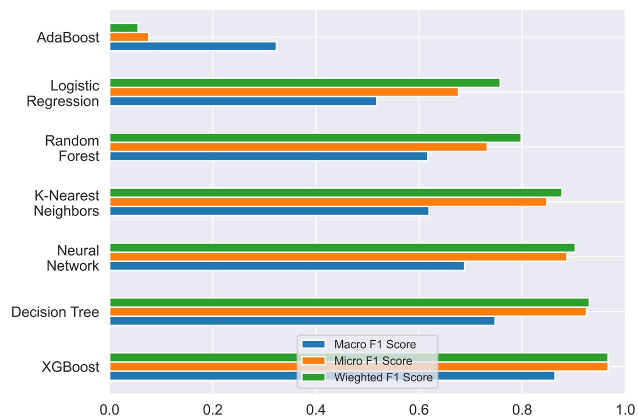

FIGURE 11. Evaluation Metrics of Different Models.

The XGBoost model demonstrates performance in predicting student qualifications for scholarships in each class, as detailed in Figure 12. Figure 13 further details this performance by visualizing the precision-recall (PR) curve. This curve illustrates the trade-off between precision, the proportion of correctly identified positive examples, and recall, the proportion of actual positives that are correctly classified [66].

The model achieves an impressive precision of 0.9813 and recall of 0.9821, reflecting highly accurate predictions for students unlikely to qualify. The nearly perfect average precision score of 0.9986 for the "No scholarship" category indicates the model's exceptional capability in accurately identifying students who do not qualify for a scholarship. This

TABLE 8. Model evaluation metrics.

Metrics	Formula	Interpretation
Precision	$P = \frac{TP}{TP+FP} \in [0, 1]$	Precision measures the proportion of examples labeled as positive that are true positives. This metric evaluates the accuracy of a model in identifying instances of a specific class. [61].
Recall	$R = \frac{TP}{TP+FN} \in [0, 1]$	Recall measures the proportion of correctly classified positive patterns [62].
F1 score	$F1 = \frac{2 \times P \times R}{P+R} \in [0, 1]$	F1 score is the harmonic mean of precision and recall, providing a single metric that balances both concerns [61].
Average precision	$AP = \sum_{t=1}^{T} (R_t - R_{t-1}) P_t \in [0, 1]$	Average precision is a widely used metric that considers both recall and precision [63].
Balanced accuracy	$BA = \frac{1}{N} \sum_{i=1}^{N} R_i \in [0, 1]$	Balanced accuracy is the arithmetic mean of class-specific accuracies. It represents the average recall for each class, making it suitable for handling imbalanced datasets.

t: Threshold
T: Number of thresholds
N: Number of classes

TABLE 9. Average scores.

Average scores	Interpretation
Macro average	Macro average computes the metric independently for each class and then takes the average [64]. It treats all classes equally, regardless of the number of instances in each class.
Micro average	Micro average aggregate the contributions of all classes to compute the average metric [64]. It treats every individual prediction equally without considering the class it belongs to. This is particularly useful when the class distribution is imbalanced, as it gives a more balanced view of the model's performance across all classes.
Weighted average	Weighted average computes the metric for each class, like macro average, but it takes the support of each class into account. This means that classes with more instances will have a greater impact on the final score. It's particularly useful when dealing with imbalanced datasets, as it provides a balanced view that considers class prevalence.

TABLE 10. XGBoost model configuration and its settings.

Parameter	booster	device	learning_rate	gamma	max_depth	min_child_weight	max_delta_step	subsample	colsample_bytree	n_estimators
Argument	gbtree	cpu	0.3	0	8	1	2	1	1	600

TABLE 11. Evaluation metrics of different models.

	Macro F1	**Micro F1**	**Weighted F1**	**mAP**[1]	**BA**[2]
AdaBoost	0.3233	0.0757	0.0555	0.4526	0.6198
Logistic Regression	0.5182	0.6767	0.7573	0.5165	0.7786
Random Forest	0.6171	0.7326	0.7982	0.7189	**0.9080**
K-Nearest Neighbors	0.6196	0.8478	0.8772	0.5950	0.7813
Neural Network (Multi-layer Perceptron)	0.6885	0.8865	0.9030	0.7256	0.8074
Decision Tree	0.7475	0.9244	0.9230	0.6258	0.8064
XGBoost	**0.8634**	**0.9666**	**0.9664**	**0.9168**	0.8675

[1] Mean average precision.
[2] Balanced accuracy.

high precision is particularly valuable given that this category likely comprises the majority of the test set, reflecting the model's effectiveness in recognizing the most common outcome. The PR curve for no scholarship class remains close to the top of the graph, indicating a high level of precision across all levels of recall. This implies that the model is consistently accurate in predicting students who do not receive a scholarship, even as it tries to capture more of the true "No scholarship" instances (higher recall).

For the "Excellent scholarship" category, the model also performs impressively, with an average precision score of 0.9875, a precision of 0.9256, and a recall of 0.9860. This high score suggests that the model is very effective at pinpointing the top-performing students who are laudable of the highest scholarship awards. Given that this category represents a smaller, more select group of students, the model's ability to distinguish these top achievers with high precision is crucial for the integrity and fairness of the scholarship allocation process. The curve of this class stays near the top, starting with very high precision and slightly declining as recall increases. This demonstrates that the model is very precise in identifying students who deserve the excellent scholarship, with minimal false positives until it reaches a high level of recall. This is indicative of strong predictive power for distinguishing the top-tier students.

However, there is a noticeable decline in the model's performance metrics such as precision, recall, and F1-score as the scholarship category lowers from Excellent to Very Good to Good. The PR curve for the very good scholarship class shows more of a gradual decline, starting from a moderate precision level. The noticeable decline as recall increases suggests that the model faces challenges in maintaining precision while trying to include more true positives for the "Very good scholarship" category. The moderate starting point for precision indicates some confusion or overlap with other categories, which might be due to less distinct features between the "Very good scholarship" and "No scholarship" categories.

The PR curve for the "Good scholarship" starts lower and shows a steeper decline than the other categories. This signifies difficulty in accurately predicting the good scholarship category, with a notable trade-off between capturing more true positives and maintaining precision. The lower and more rapidly declining curve suggests that the model struggles with a higher rate of false positives in this category.

The observed trend suggests that scholarships are distributed sequentially due to the limited number of awards available. Within each class, scholarships are allocated from the highest achievers downward, categorized according to each student's academic achievements and extracurricular contributions. This method allows the model to excel in accurately identifying the most praiseworthy students, even if they belong to the least populous category. As students' academic and extracurricular points decline, so does their likelihood of receiving a scholarship, leading to an increase in misclassifications particularly in the "Very good scholarship" and "Good scholarship" to "No scholarship" categories. Nevertheless, the model effectively prioritizes the most qualified students, resulting in those who excel being duly recognized and rewarded.

FIGURE 12. XGBoost model evaluation for each class.

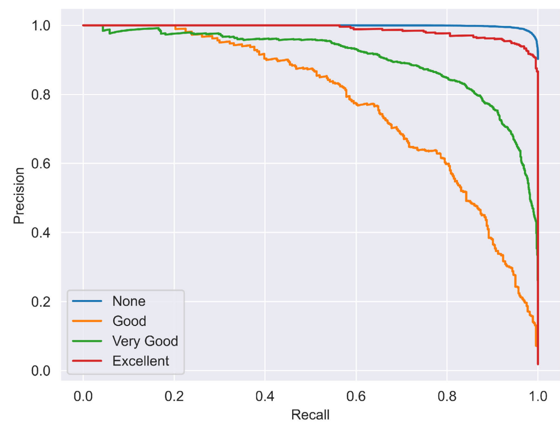

FIGURE 13. Precision recall curve.

VI. DISCUSSION

This study presents a transformative approach to scholarship management through the deployment of a blockchain-based platform. By integrating zk-rollup scaling solutions, the platform achieves enhanced scalability and inherits the robust security of L1 blockchains. Our empirical results demonstrate the platform's capability to reduce transaction processing times by over 60% and transaction costs by nearly 90%, marking a substantial improvement over EVM-based solutions. Furthermore, integrating an ML-based recommendation system for scholarship qualifications within a blockchain-based scholarship management platform can significantly motivate students. With a balanced accuracy rate of 86.75%, the XGBoost model serves as a reliable recommendation system, prioritizing truly admirable students and helping them avoid missing out on crucial financial support opportunities.

In terms of real-world applicability, this system stands as a testament to the power of combining blockchain technology and ML to create a fairer, more efficient, and transparent scholarship management system. Let us delve into the benefits that this platform offers to involved stakeholders.

For students, it offers increased access to relevant scholarship opportunities through recommendations, enhanced privacy protection through SSI and VCs, improved motivation and goal-setting through data-driven feedback, reduced application complexity, and fair consideration based on merit. Academic institutions benefit from streamlined scholarship management processes, improved ability to match students with appropriate scholarships, and enhanced reputation for supporting student success.

Scholarship providers and donors gain increased transparency in fund allocation and distribution, enhanced trust through immutable transaction records, reduced administrative costs, improved ability to target specific student demographics or academic achievements, and greater assurance that funds are reaching deserving students. Government and regulatory bodies benefit from improved oversight and audit capabilities, reduced potential for fraud, and enhanced ability to analyze scholarship distribution patterns for policy-making. Technology providers find new market opportunities and potential for cross-sector collaborations, while society at large benefits from increased educational opportunities, potential reduction in student debt burden, and enhanced meritocracy in educational funding distribution.

Blockchain technology guarantees transparency and trust by enabling smart contracts that define clear fund distribution rules. These contracts render all financial transactions traceable and immutable, boosting donor confidence. Additionally, zk-rollup technology minimizes transaction fees and processing times, encouraging greater donor participation. Students often hesitate to share sensitive academic and personal information with external organizations. SSI empowers students to control their data by sharing only sensitive documents with their academic institution. These institutions can then issue VCs, which students can use to generate ZKPs for scholarship applications, providing the necessary information without disclosing sensitive details. The system's machine learning component identifies potential scholarship recipients, mitigating the risk of overlooking deserving candidates. By analyzing academic data and recommending suitable scholarships, the system personalizes the journey, fostering goal setting through predicted qualification benchmarks. Students receive data-driven feedback on their progress, boosting confidence and a sense of achievement as they climb toward scholarship qualifications.

This approach represents a leap forward in scholarship management, addressing long-standing issues of inefficiency, opacity, and inequity. By leveraging the strengths of blockchain technology and ML, it creates a more inclusive, transparent, and effective system that benefits all stakeholders in the educational ecosystem. As this technology matures and gains wider adoption, it has the potential to revolutionize how we approach educational funding and support, ultimately contributing to a more educated and equitable society.

A. POTENTIAL LIMITATIONS AND CHALLENGES IN IMPLEMENTATION

The proposed platform, leveraging blockchain technology, SSI, zk-rollups, and ML, presents a promising solution for revolutionizing scholarship management. However, several challenges must be addressed for successful implementation and widespread adoption.

1) TRADE-OFF BETWEEN SCALABILITY AND SECURITY

The balance between scalability and security is a critical issue that will likely influence the implementation of blockchains in the field where they are used. As systems grow to accommodate increasingly complex and resource-intensive users, transactions, or data volumes, upholding strong security becomes more difficult and time-consuming. Scalability can introduce vulnerabilities that expose the system to attacks. However, strict security measures can have an impact on system performance and scalability. Balancing these two aspects requires a careful evaluation of priorities, to increase scalability without risking security, or enhancing security without severely limiting the system's ability to grow and adapt to new demands.

The EVM operates within a decentralized network of nodes that independently verify smart contract executions and state changes. These nodes process transactions and achieve consensus through Ethereum's consensus mechanism, guaranteeing all nodes maintain a consistent and agreed-upon state of the blockchain. This on-chain execution enforces that all smart contract operations comply with network rules, preserving decentralization but slowing down computation. Each block has limited space for transaction data and gas, restricting computational capacity per block. These constraints limit Ethereum's computational scalability over time.

A potential solution to enhance scalability lies in decoupling contract computation from Ethereum's state computation. However, this introduces the challenge of maintaining the EVM's rule enforcement when computation occurs off-chain. zkEVMs present an innovative solution to this challenge by combining ZKPs with the EVM. zkEVMs utilize ZKPs to verify the validity of off-chain computations, enforcing adherence to EVM rules. These proofs allow the Ethereum chain's state to be updated with minimal, yet verifiable data, significantly reducing the on-chain computational burden. Consequently, zkEVMs offer unparalleled scalability to Ethereum applications, alongside cost reductions and robust security and decentralization. Our solution leverages the benefits of zkEVMs to strengthen scalability and privacy while upholding the core principles of security and decentralization in the Ethereum ecosystem.

2) BLOCKCHAIN INTEROPERABILITY

Blockchains often struggle to scale to the volumes required for widespread adoption. Programmable blockchains address this by relocating users, funds, and data to parallel structures such as sidechains and rollups (including our platform's implementation of zk-rollups). Yet, this solution presents a new issue: it segments the user experience and disrupts interoperability. This arises because donors may hold assets across various blockchains, and recipients may prefer to receive funds in different cryptocurrencies or on other blockchains than the one on which the platform operates.

Furthermore, the rapid evolution of the blockchain ecosystem presents an additional challenge. As new blockchain networks proliferate, decentralized applications like our SSM platform face constant pressure to adapt and expand compatibility. Failure to keep pace with these developments risks limiting the platform's reach and excluding potential donors who operate on newer or alternative blockchain networks.

Chain Abstraction appears as a promising solution to these challenges. This approach allows donors to contribute to scholarship campaigns regardless of their native blockchain or the specific tokens they hold.

3) TRUST IN SELF-SOVEREIGN IDENTITY

The core concept of SSI hinges on the trust established between the platform and academic institutions. The platform must have confidence in the academic institution's reputation and reliability. While cryptographic techniques confirm that a VC has not been tampered with and was issued by the stated academic institution, they do not validate the accuracy of the underlying information. Thus, balancing decentralization with necessary governance is a key challenge. Although SSI aims to minimize reliance on central authorities, some level of governance and oversight is still required to ensure that academic institutions are trustworthy.

4) REGULATORY HURDLE

Regulatory compliance poses a hurdle, as different countries have varying regulations regarding blockchain use. To overcome this, the platform should be designed with flexibility to adapt to different regulatory frameworks, and efforts should be made to engage with policymakers to create favorable regulations for educational blockchain applications.

5) USER ADOPTION AND TECHNOLOGICAL BARRIERS

User adoption is another critical challenge, as students, educational institutions, and donors may struggle with the new method due to unfamiliarity with blockchain technology. The blockchain technology underpinning SSM, while powerful, can be intimidating to those without a technical background. Moreover, the implementation of SSI adds another layer of complexity that may deter users. This can be mitigated by creating an intuitive, user-friendly interface that abstracts complex blockchain operations and developing detailed onboarding materials.

B. XGBOOST'S EFFICIENCY

Before closing this section, we will discuss why XGBoost "wins" this competition among several models, as detailed in Table 11. This success is attributed to two main factors: the quality of the dataset and the effectiveness of the XGBoost algorithm.

Firstly, regarding data quality, our dataset comprises a diverse set of features related to students, as outlined in Table 6. The data was collected from VNUHCM - UIT spanning from 2013 to 2022. Below, we provide our data quality analysis, evaluating five key attributes: Accuracy, Completeness, Consistency, Validity, and Uniqueness.

a: ACCURACY

The dataset has undergone a review focusing on various aspects of data quality to validate its readiness and reliability before model training. The following checks were conducted:

- *Semester GPA range*: All GPA values are within the expected 0 to 10 scale, indicating no issues with the GPA range.
- *Scholarship Range*: Scholarship qualification values are correctly within the 0 to 3 range, adhering to the predefined categorization.
- *GPA*: The dataset correctly excludes students with a GPA less than 7 from scholarship qualifications. This assures that only students who achieve sufficiently high academic results will be considered for scholarships
- *Total credit*: The dataset appropriately omits students from scholarship consideration if their total_credit is below 14. This is because having a total_credit less than 14 signifies a workload below the standard threshold for a student within a semester.
- *Failed course impact*: Students marked as having failed a course (has_failed_course is True) in a certain period of academic assessment are appropriately disqualified for scholarships during that timeframe, upholding adherence to the scholarship rules.
- *Missing Values*: The dataset contains no missing values across all columns, indicating high completeness.

b: COMPLETENESS

This complete data presence allows for analyses conducted on this dataset to proceed without being hindered by gaps or the need for imputation techniques to address missing data. The dataset exhibits a 100% completeness rate, with no missing values across all entries and columns.

c: CONSISTENCY

The Pearson correlation coefficient between **GPA** and **Extracurricular point** is around 0.62, signifying a moderate to strong positive correlation. This implies that, in most cases, an increase in students' GPA is associated with an increase in their extracurricular points.

The scatter plot in Figure 14 displays a positive relationship between GPA and extracurricular points. The distribution of points suggests variability in how extracurricular points are awarded relative to GPA. A dense clustering of data points at the higher end of the GPA scale indicates that students with higher GPAs also engage significantly in extracurricular activities. Therefore, it can be observed that with the rise in GPA, extracurricular points generally increase as well.

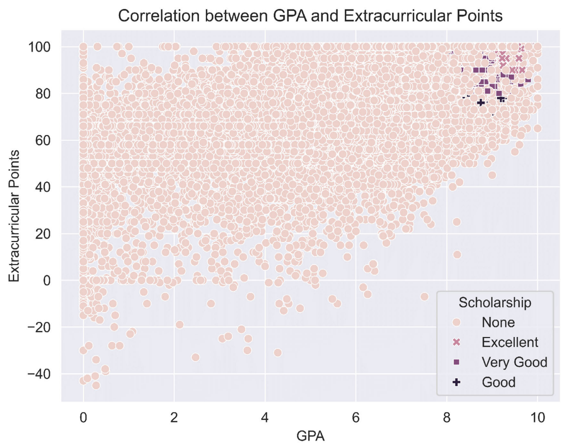

FIGURE 14. Correlation between GPA and Extracurricular Points.

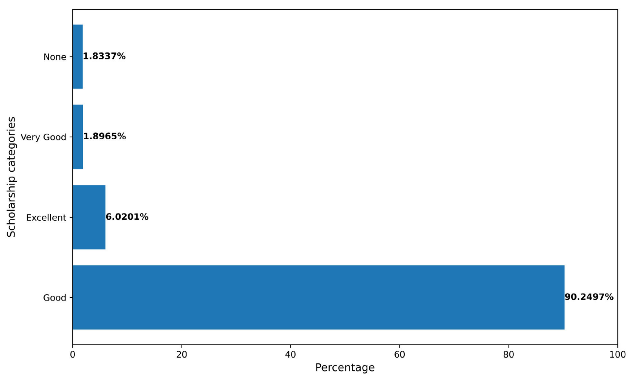

FIGURE 15. Scholarship distribution.

d: VALIDITY

Figure 15 shows the distribution of scholarship in the dataset. It reveals that nearly 90.25% of students do not receive any scholarship. This aligns with the competitive nature of scholarship awards and indicates that scholarships are reserved for a relatively small percentage of the student population. The dataset's scholarship policy stipulates that approximately 10% of students are to receive scholarships.

e: UNIQUENESS

When examining the dataset, it is noted that the combination of student ID, semester, and school year uniquely identifies all 78,039 records. This indicates that every record uniquely reflects a distinct moment in a student's academic path, with no repetition concerning these combined attributes. This uniqueness is essential for precisely monitoring and analyzing student advancement through time. Other columns don't need to maintain uniqueness.

Moving on to the second factor, the robustness of the XGBoost algorithm significantly contributes to its success. Tree boosting is generally effective because it fits additive tree models with high representational ability using adaptively determined neighborhoods. This adaptive capability allows the model to apply varying degrees of flexibility in different regions of the input space. Consequently, XGBoost can perform automatic feature selection and capture high-order interactions without breaking down, making it robust to the curse of dimensionality and capable of balancing the bias-variance trade-off effectively [67].

The Gradient Boosting Machine (GBM) was a dominant tree-boosting method for many years. However, XGBoost has emerged as a powerful contender, winning numerous ML competitions. In 2015, 17 out of 29 winning solutions on Kaggle utilized XGBoost. Similarly, XGBoost dominated the prestigious KDD Cup 2015, with all top 10 solutions employing it.

XGBoost utilizes a different form of boosting technique from GBM. While GBM uses gradient boosting, interpreted as gradient descent in function space, XGBoost's boosting algorithm can be viewed as Newton's method in function space. This involves a higher-order approximation at each iteration, allowing XGBoost to learn more effective tree structures. XGBoost also offers additional parameters not found in GBM, most notably the penalization of individual trees in the additive tree model. These parameters influence both tree structure and leaf weights, reducing variance within each tree. Additionally, XGBoost includes a randomization parameter to decorrelate individual trees, further reducing the overall variance of the additive tree model [67].

C. FUTURE WORK

In future work, we aim to address several challenges associated with the proposed platform and expand its capabilities:

In response to the interoperability challenge, we plan to enhance our platform by integrating the concept of chain abstraction. This enhancement will enable users to contribute or receive scholarship funds using their chosen cryptocurrency, all within a unified user interface, thus maintaining a consistent user experience without navigating away from the platform's interface.

While our current ML model focuses on AES prevalent in Vietnamese higher education institutions, we recognize the need for a more comprehensive approach. We intend to broaden our data collection efforts appreciably, incorporating a diverse range of features such as family background, socioeconomic factors, and career aspirations. Our goal is to build and promote a standardized, globally relevant dataset. This enriched data will enable us to develop more sophisticated ML modules capable of supporting a wide variety of scholarships, from merit-based to need-based and specialized fields of study.

SSI places a weighty burden on academic institutions to verify student documents for VC issuance. To address this growing administrative challenge, Neural Networks have emerged as a powerful solution. Stacked Intelligent

Metasurfaces [68], particularly hardware type III, show potential for implementing Deep Neural Network functionalities, which can efficiently manage complex tasks like image recognition. Additionally, Convolutional Neural Networks [69] provide a promising approach for image recognition. These technologies can streamline the processing and verification of student documents, such as academic transcripts, standardized test scores, and extracurricular records. These solutions could potentially enhance the efficiency of VC issuance by reducing the time and effort required to verify student data.

VII. CONCLUSION

Qualified students should never be excluded from the scholarship opportunities they deserve. Scholarships are crucial, offering vital financial support that covers tuition costs and enables learners to seek higher education free from the shackles of financial debt. Being awarded a scholarship also greatly bolsters a student's confidence and self-worth, recognizing and validating their talents and prospects.

Our work focuses on developing a decentralized scholarship management system that leverages blockchain technology to warrant security, transparency, and decentralization, while effectively addressing the scalability and privacy concerns present in existing related work. SSM overcoming scalability challenges faced by integrating zkEVMs, SSM achieves over 60% faster transaction processing times and reduces transaction fees by approximately 90%. To preserve privacy, SSM utilizes ZKPs and SSI to safeguard smart contract execution and user data. Furthermore, SSM leverages the power of ML to prevent qualified students from missing out on scholarship opportunities. SSM optimizes the management process and establishes a more efficient and effective system by utilizing an XGBoost ML model with a nearly 92% mean average precision in predicting student scholarship qualifications.

ACKNOWLEDGMENT

This research was supported by The VNUHCM-University of Information Technology's Scientific Research Support Fund.

REFERENCES

[1] M. A. Durand, "Scholarship program," *J. SMPTE*, vol. 80, no. 1, pp. 45–46, Jan. 1971, doi: 10.5594/J13556.

[2] L. Yua, "Practice and investigation of college student funded work," *Chin. Foreign Med. Res.*, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:76105143

[3] T. Osterland and T. Rose, "From a use case categorization scheme towards a maturity model for engineering distributed ledgers," in *Blockchain and Distributed Ledger Technology Use Cases: Applications and Lessons Learned*, H. Treiblmaier and T. Clohessy, Eds. Cham, Switzerland: Springer, 2020, pp. 33–50, doi: 10.1007/978-3-030-44337-5_2.

[4] T. Lavaur, J. Detchart, J. Lacan, and C. P. C. Chanel, "Modular zk-rollup on-demand," *J. Netw. Comput. Appl.*, vol. 217, Aug. 2023, Art. no. 103678.

[5] R. Soltani, U. T. Nguyen, and A. An, "A survey of self-sovereign identity ecosystem," *Secur. Commun. Netw.*, vol. 2021, pp. 1–26, Jul. 2021.

[6] *1.1 Billion 'Invisible' People Without Id Are Priority for New High Level Advisory Council on Identification for Development*, World Bank, Washington, DC, USA, 2017.

[7] C. Satchell, G. Shanks, S. Howard, and J. Murphy, "Identity crisis: User perspectives on multiplicity and control in federated identity management," *Behaviour Inf. Technol.*, vol. 30, no. 1, pp. 51–62, Jan. 2011.

[8] S. Kayikci and T. M. Khoshgoftaar, "Blockchain meets machine learning: A survey," *J. Big Data*, vol. 11, no. 1, p. 9, Jan. 2024, doi: 10.1186/s40537-023-00852-y.

[9] S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh, and W.-C. Hong, "Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward," *IEEE Access*, vol. 8, pp. 474–488, 2020.

[10] M. B. Fuller, "A history of financial aid to students," *J. Student Financial Aid*, vol. 44, no. 1, p. 4, Jul. 2014.

[11] M. Dachyar, M. Salman, and R. Nurcahyo, "Strategies to improve the education and research scholarship program at the universities," *TEM J.*, vol. 12, pp. 389–395, Feb. 2023, doi: 10.18421/TEM121-48.

[12] M. Nofer, P. Gomber, O. Hinz, and D. Schiereck, "Blockchain," *Bus. Inf. Syst. Eng.*, vol. 59, no. 3, pp. 183–187, Jun. 2017, doi: 10.1007/s12599-017-0467-3.

[13] T. Nguyen and T. N. Gia, *Novel Smart Homecare IoT System With Edge-AI and Blockchain*. Singapore: Springer Nature, 2023, pp. 293–317.

[14] T. Nguyen, H. Nguyen, J. Partala, and S. Pirttikangas, "Trusted-MaaS: Transforming trust and transparency mobility-as-a-service with blockchain," *Future Gener. Comput. Syst.*, vol. 149, pp. 606–621, Dec. 2023.

[15] T. Nguyen, R. Katila, and T. N. Gia, "An advanced Internet-of-Drones system with blockchain for improving quality of service of search and rescue: A feasibility study," *Future Gener. Comput. Syst.*, vol. 140, pp. 36–52, Mar. 2023.

[16] M. Di Pierro, "What is the blockchain?" *Comput. Sci. Eng.*, vol. 19, no. 5, pp. 92–95, 2017.

[17] D. Yaga, P. Mell, N. Roby, and K. Scarfone, "Blockchain technology overview," 2019, *arXiv:1906.11078*.

[18] N. Szabo, "Formalizing and securing relationships on public networks," *1st Monday*, vol. 2, no. 9, Sep. 1997, doi: 10.5210/fm.v2i9.548. [Online]. Available: https://firstmonday.org/ojs/index.php/fm/article/view/548

[19] W. Zou, D. Lo, P. S. Kochhar, X. D. Le, X. Xia, Y. Feng, Z. Chen, and B. Xu, "Smart contract development: Challenges and opportunities," *IEEE Trans. Softw. Eng.*, vol. 47, no. 10, pp. 2084–2106, Oct. 2021.

[20] Z. Zheng, S. Xie, H.-N. Dai, W. Chen, X. Chen, J. Weng, and M. Imran, "An overview on smart contracts: Challenges, advances and platforms," *Future Gener. Comput. Syst.*, vol. 105, pp. 475–491, Apr. 2020.

[21] S. Rouhani and R. Deters, "Security, performance, and applications of smart contracts: A systematic survey," *IEEE Access*, vol. 7, pp. 50759–50779, 2019.

[22] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof-systems," in *Proc. Providing Sound Found. Cryptogr., Work Shafi Goldwasser Silvio Micali*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 203–225, doi: 10.1145/3335741.3335750.

[23] M. Blum, P. Feldman, and S. Micali, "Non-interactive zero-knowledge and its applications," in *Proc. Providing Sound Found. Cryptography, Work Shafi Goldwasser Silvio Micali*, 2019, pp. 329–349.

[24] T. Chen, H. Lu, T. Kunpittaya, and A. Luo, "A review of zk-SNARKs," 2022, *arXiv:2202.06877*.

[25] J. Groth and A. Sahai, "Efficient non-interactive proof systems for bilinear groups," in *Advances in Cryptology—EUROCRYPT 2008*. Berlin, Germany: Springer, Apr. 2008, pp. 415–432.

[26] M. Maller, S. Bowe, M. Kohlweiss, and S. Meiklejohn, "Sonic: Zero-knowledge SNARKs from linear-size universal and updatable structured reference strings," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2019, pp. 2111–2128.

[27] L. Zhang, H. Xu, O. Onireti, M. A. Imran, and B. Cao, "How much communication resource is needed to run a wireless blockchain network?" *IEEE Netw.*, vol. 36, no. 1, pp. 128–135, Jan. 2022.

[28] Y. Gong, Y. Jin, Y. Li, Z. Liu, and Z. Zhu, "Analysis and comparison of the main zero-knowledge proof scheme," in *Proc. Int. Conf. Big Data, Inf. Comput. Netw. (BDICN)*, Jan. 2022, pp. 366–372.

[29] F. Schardong and R. Custódio, "Self-sovereign identity: A systematic review, mapping and taxonomy," *Sensors*, vol. 22, no. 15, p. 5641, Jul. 2022.

[30] Š. Čučko, V. Keršič, and M. Turkanović, "Towards a catalogue of self-sovereign identity design patterns," *Appl. Sci.*, vol. 13, no. 9, p. 5395, Apr. 2023.

[31] D. Reed, M. Sporny, D. Longley, C. Allen, R. Grant, M. Sabadello, and J. Holt, "Decentralized identifiers (dids) v1. 0," W3C, Cambridge, MA, USA, Draft Community Group Rep., 2020. Accessed: May 10, 2024. [Online]. Available: https://www.w3.org/TR/did-core/

[32] WWW Consortium. (2019). *Verifiable Credentials Data Model 1.0: Expressing Verifiable Information on the web*. [Online]. Available: https://www.w3.org/TR/vc-data-model/?#core-data-model

[33] E. Krul, H.-Y. Paik, S. Ruj, and S. S. Kanhere, "SoK: Trusting self-sovereign identity," 2024, *arXiv:2404.06729*.

[34] A. Satybaldy, A. Subedi, and M. Nowostawski, "A framework for online document verification using self-sovereign identity technology," *Sensors*, vol. 22, no. 21, p. 8408, Nov. 2022.

[35] R. Hegde and S. Madival, "A review on data mining and machine learning methods for student scholarship prediction," in *Proc. 5th Int. Conf. Comput. Methodologies Commun. (ICCMC)*, Apr. 2021, pp. 923–927.

[36] K. O. Asamoah, A. P. Darko, C. O. Antwi, S. L. Kodjiku, E. S. E. B. Aggrey, Q. Wang, and J. Zhu, "A blockchain-based crowdsourcing loan platform for funding higher education in developing countries," *IEEE Access*, vol. 11, pp. 24162–24174, 2023.

[37] S. Govindarajan, S. S. Rajkumar, A. Mayan J, and A. M. Posonia, "Blockchain fundraising and charity platform," in *Proc. 2nd Int. Conf. Vis. Towards Emerg. Trends Commun. Netw. Technol. (ViTECoN)*, May 2023, pp. 1–6.

[38] A. O. A. K. Alassaf and F. H. Yusoff, "Multi-point fundraising and distribution via blockchain," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, 2021, doi: 10.14569/IJACSA.2021.0120755. [Online]. Available: https://thesai.org/Publications/ViewPaper?Volume=12&Issue=7&Code=IJACSA&SerialNo=55

[39] R. Dange, A. Sawant, A. Chavan, P. Bhardwaj, and A. Bundele, "Decentralized fundraising application using blockchain," in *Proc. IEEE Int. Conf. Blockchain Distrib. Syst. Secur. (ICBDS)*, Sep. 2022, pp. 1–5.

[40] B. Hu and H. Li, "Research on charity system based on blockchain," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 768, no. 7, Aug. 2020, Art. no. 072020, doi: 10.1088/1757-899X/768/7/072020.

[41] J. Anju and R. Gini, *Charity System Based Using Blockchain*, Aug. 2023, doi: 10.59544/PREY8262/NGCESI23P8.

[42] P. Mhatre, T. Patil, S. Thorat, and K. Thorat, "Donation based system using blockchain," *Int. J. Adv. Res. Sci., Commun. Technol.*, pp. 394–396, Apr. 2023, doi: 10.48175/IJARSCT-9591.

[43] H. Saleh, S. Avdoshin, and A. Dzhonov, "Platform for tracking donations of charitable foundations based on blockchain technology," in *Proc. Actual Problems Syst. Softw. Eng. (APSSE)*, Nov. 2019, pp. 182–187.

[44] M. Darshan, S. Raswanth, S. V. Akella, and P. Kumar, "A secured distributed ledger based fundraising framework using smart contracts," in *Proc. IEEE 4th Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Sep. 2021, pp. 1–5.

[45] Y. Zhou, "Understanding users' reaction to blockchain technology on the online fundraising platform—Evidence from scenario simulation experiments," in *Proc. Int. Conf. Comput. Inf. Sci. Artif. Intell. (CISAI)*, Sep. 2021, pp. 301–305.

[46] M. A. Rashid, K. Deo, D. Prasad, K. Singh, S. Chand, and M. Assaf, "TEduChain: A blockchain-based platform for crowdfunding tertiary education," *Knowl. Eng. Rev.*, vol. 35, p. e27, 2020, doi: 10.1017/S0269888920000326.

[47] U. Tekguc, A. Adalier, and K. Yurtkan, "Scholarchain: The scholarship management platform with blockchain and smart contracts technology," in *Proc. Eurasia Educ. Social Sci.*, vol. 18, 2020, pp. 86–91.

[48] P. Bedi, P. Gole, S. Dhiman, and N. Gupta, "Smart contract based central sector scheme of scholarship for college and university students," *Proc. Comput. Sci.*, vol. 171, pp. 790–799, Sep. 2020.

[49] S. Jadhav, N. Pise, and P. Karwa, "Scholar block: Enhancing corporate social responsibility through a blockchain-based scholarship system," in *Proc. 2nd Int. Conf. Informat. (ICI)*, vol. 70, Nov. 2023, pp. 1–6.

[50] J. Swati and P. Nitin, "CryptoScholarChain: Revolutionizing scholarship management framework with blockchain technology," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, Aug. 2024. [Online]. Available: https://www.proquest.com/scholarly-journals/cryptoscholarchain-revolutionizing-scholarship/docview/2869803985/se-2?accountid=27468

[51] N. C. Hoang, P. T. Hua, T. Nguyen, K. Tan-Vo, T.-A. Nguyen-Hoang, T. Nguyen, and N.-T. Dinh, "SSSM: A secure and scalable approach for scholarship funding management based on blockchain technology with zk-rollups," in *Intelligent Systems Design and Applications*, A. Abraham, S. Pllana, T. Hanne, and P. Siarry, Eds. Cham, Switzerland: Springer, 2024, pp. 341–350.

[52] W. Ilham, T. E. Putri, P. Sokibi, and Kusnadi, "Application of fuzzy multy attribute decision-making method in decision-making system for determining the provision of achievement scholarship in SMPN 1 simpati," *J. Phys., Conf. Ser.*, vol. 1641, no. 1, Nov. 2020, Art. no. 012052.

[53] W. D. Ahmad and A. Abu Bakar, "Classification models for higher learning scholarship award decisions," *Asia-Pacific J. Inf. Technol. Multimedia*, vol. 7, no. 2, pp. 131–145, Dec. 2018.

[54] B. Arifitama, "Decision support system scholarship selection using simple additive weighting (SAW) method," *JISA(Jurnal Informatika dan Sains)*, vol. 5, no. 1, pp. 80–84, Jun. 2022.

[55] V. Buterin, "Ethereum: platform review," in *Opportunities and Challenges for Private and Consortium Blockchains*, vol. 45, 2016.

[56] T. Bayan and R. Banach, "Exploring the privacy concerns in permissionless blockchain networks and potential solutions," in *Proc. IEEE Int. Conf. Smart Inf. Syst. Technol. (SIST)*, May 2023, pp. 567–572.

[57] S. Park, S. Oh, and H. Kim, "Performance analysis of DAG-based cryptocurrency," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2019, pp. 1–6.

[58] A. Donmez and A. Karaivanov, "Transaction fee economics in the ethereum blockchain," *Econ. Inquiry*, vol. 60, no. 1, pp. 265–292, Jan. 2022.

[59] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[60] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *J. Mach. Learn. Res.*, vol. 1, pp. 113–141, Dec. 2000.

[61] N. Japkowicz, *Assessment Metrics for Imbalanced Learning*. Hoboken, NJ, USA: Wiley, 2013, ch. 8, pp. 187–206.

[62] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 1–11, Mar. 2015.

[63] M. Zhu, "Recall, precision and average precision," Dept. Statist. Actuarial Sci., Univ. Waterloo, Waterloo, ON, USA, 2004, vol. 2, no. 30, p. 6. Accessed: Apr. 21, 2024. [Online]. Available: https://datascience-intro.github.io/1MS041-2022/Files/AveragePrecision.pdf

[64] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, "Boosting methods for multi-class imbalanced data classification: An experimental review," *J. Big Data*, vol. 7, no. 1, pp. 1–47, Dec. 2020.

[65] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, vol. 11, New York, NY, USA. Association for Computing Machinery, Aug. 2016, pp. 785–794.

[66] K. Boyd, V. S. Costa, J. Davis, and C. D. Page, "Unachievable region in precision-recall space and its effect on empirical evaluation," in *Proc. Int. Conf. Mach. Learn.*, Dec. 2012, p. 349.

[67] D. Nielsen, "Tree boosting with Xgboost-why does Xgboost win 'every' machine learning competition?" Master's thesis, NTNU, Trondheim, Norway, 2016.

[68] H. Liu, J. An, X. Jia, S. Lin, X. Yao, L. Gan, B. Clerckx, C. Yuen, M. Bennis, and M. Debbah, "Stacked intelligent metasurfaces for wireless sensing and communication: Applications and challenges," 2024, *arXiv:2407.03566*.

[69] X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. Nguyen, S. Chu, B. Little, D. Hicks, R. Morandotti, A. Mitchell, and D. Moss, "11 tera-flop per second photonic convolutional accelerator for deep learning optical neural networks," *Preprints*, Nov. 2020.

TU-ANH NGUYEN-HOANG (Member, IEEE) received the Ph.D. degree from the Mathematical Foundations of Computer Science and Computational Systems, University of Science, VNU-HCM. She is currently an Associate Professor with the University of Information Technology, Vietnam National University, Ho Chi Minh City (VNU-HCM). Her research interests include data mining, soft computing, and machine learning. She is actively involved in numerous research projects and has a prolific publication record in international conferences and journals. Her work significantly contributes to advancing these specialized fields.

NGOC CU HOANG is currently pursuing the Bachelor of Information Technology degree with the University of Information Technology. He is currently a Member Institution of Vietnam National University of Ho Chi Minh City (VNUHCM). His research interest includes blockchain technology, he is dedicated to leveraging the technology for practical solutions.

PHU THIEN HUA is currently pursuing the bachelor's degree in information technology with the University of Information Technology, Vietnam National University—Ho Chi Minh City (VNU-HCM). His research interest includes blockchain technology, areas in which he has contributed to several projects and published papers in international conferences. His dedication to this emerging technology is driven by a passion for innovation and a desire to contribute to the development of secure and decentralized systems.

MONG-THY NGUYEN THI received the Master of Arts degree in teaching English as a second language (TESL) from Benedictine University, USA. She is currently an Educator with the Center for Foreign Languages, University of Information Technology—Vietnam National University, Ho Chi Minh City (VNU-HCM). Her research interest includes English language teaching (ELT) methodology. Her passionate about improving language education, she contributes significantly to advancing ELT practices at her institution.

THU-THUY TA received the Master of Science degree in computer science from the University of Information Technology, Vietnam National University—Ho Chi Minh City (VNU-HCM). She is currently a Lecturer with the Faculty of Information Science and Engineering, VNU-HCM. Her research interests include data science and machine learning, areas in which she has contributed to several projects and published papers in international conferences and journals. Her work reflects a commitment to advancing the fields of data science and machine learning.

THU NGUYEN received the Master of Science degree in computer science from the University of Information Technology, Vietnam National University—Ho Chi Minh City (VNU-HCM). She is currently a Lecturer with the Faculty of Information Science and Engineering, VNU-HCM. Her research interests include data science, AI, and machine learning. She has participated in numerous projects and published papers in prestigious international conferences and journals, demonstrating her commitment to advancing these fields.

KHOA TAN-VO received the Master of Science degree in information technology from the University of Information Technology, Vietnam National University—Ho Chi Minh City (VNU-HCM). He is currently a Lecturer with the Faculty of Information Science and Engineering, VNU-HCM. Specializing in software engineering and blockchain technology, he actively contributes to various research projects and has a robust portfolio of publications in prestigious international conferences and journals. His dedication to advancing technology and education underscores his commitment to excellence in his field.

NGOC-THANH DINH received the Ph.D. degree in information and telecommunication engineering from Soongsil University. He was an Assistant Professor at Soongsil University, South Korea. He is a Lecturer at SFA, Industrial University of Ho Chi Minh City (IUH), Vietnam. His current research interests include the Internet of Things and cloud computing, 5G-6G networking, next-generation networks, ICT convergence, and blockchain.

HONG-TRI NGUYEN received the B.S. degree in computer science from the University of Information Technology—Vietnam National University, Vietnam, in 2015, the M.S. degree in computer science from the University of Pisa, Italy, in 2018, and the Ph.D. degree from the University of Oulu, in 2023. Recently, he has been a Postdoctoral Researcher with Aalto University. His research interests include distributed systems, blockchain technology, and information security.