PRECOGNITO- Blockchain & Machine Learning-Based Student Record Authentication System

Sheikh Mohammad Tahsin Murtaj, Mohammed Mohammed, Abu Bakar Yusif and Norzariyah Binti Yahya[a)]

Department of Computer Science, Kulliyah of Information and Communication Technology, International Islamic University Malaysia, Jln Gombak, 53100 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur

[a)]Corresponding author email: norzariyah@iium.edu.my

Abstract. Education is advancing more quickly than ever before as a result of technological advancements. Two issues brought on by technological advancements are fake transcripts and fake degrees, both of which are mostly due to data security flaws. Despite technological advancement, many domains have not been sufficiently investigated, and there will always be room for improvement. This project aims to create a tool that combines two cutting-edge technologies, machine learning (ML) and blockchain, to combat problems like degree and transcript forgery. By storing student data on the Blockchain and leveraging machine learning techniques for precise analysis, the technology can prevent additional fraud and uncertainty in student achievements and can enable accurate prediction of future employment opportunities for graduates. Machine learning algorithms will be used to train and make accurate predictions, and the requisite data will be retrieved from a Blockchain ledger. PRECOGNITO will equip the institution with a decentralized and immutable alumni database that contains verified and transparent academic records. Furthermore, this system gives employers a place to check the legitimacy of their employees' academic credentials. Moreover, PRECOGNITO allows students to upload their academic credentials to social media and professional networking sites like LinkedIn. With this system, recruiters may easily locate verified student information.

INTRODUCTION

In the swiftly evolving landscape of today's automation-driven society, numerous sectors are eagerly embracing emerging technologies to reap their manifold benefits. One industry that stands out in its enthusiastic adoption of technological progress is education. Educational institutions worldwide are actively integrating innovative tools and solutions that have demonstrated profound advantages for the holistic development of their students. Among these transformative technologies, blockchain and machine learning have emerged as promising frontiers, showcasing their potential to revolutionize education as they have in other domains.

The application of blockchain technology within the education sector has garnered substantial attention. Notably, research by [1] highlights the education sector's utilization of blockchain to establish a global assessment platform, effectively storing and managing invaluable degree information. A key attribute of blockchain technology that makes it an ideal candidate for this purpose is its inherent data immutability [2]. This feature has already found successful application across various sectors for security enhancement. It is, therefore, evident that blockchain offers a secure repository for safeguarding diplomas and certificates of academic achievement. Furthermore, in line with [3], universities and educational institutions are actively exploring systems to maintain students' academic records on the blockchain, enabling students to access and share their credentials seamlessly with potential employers via their online curriculum vitae.

Moreover, the significance of blockchain technology in education extends beyond data security. It offers the prospect of creating a safer educational environment, leveraging the latest advancements in blockchain technology. Beyond blockchain, the integration of Artificial Intelligence (AI) presents yet another exciting prospect to enhance the education sector substantially. As elucidated by [4], machine learning (ML) has evolved as a distinct area of study within computer science. Rooted in the domains of pattern recognition and computational learning theory within artificial intelligence, machine learning algorithms employ recursive learning processes to glean insights from data. This iterative approach empowers computers to conquer tasks that may appear daunting or even insurmountable to humans, bolstering their confidence in decision-making through knowledge distilled from past computations [5].

Problem Statement

The heart of the matter lies in the multifaceted challenges that confront the education sector. Firstly, the rampant proliferation of fraudulent degrees and the pervasive practice of credential forgery cast a shadow over the integrity of higher education [6]. Educational institutions are faced with the urgent need for a standardized, secure, and decentralized database capable of tracking their alumni. This database must be safeguarded against unauthorized modification or deletion to preserve

the system's integrity. Post-graduation, after successfully completing all requisite tests, students' academic records must find their place within this distributed database.

Moreover, while machine learning (ML) holds immense promise for the education sector, it faces critical pitfalls when the data it relies upon is tainted or unverified. The well-known principle of "garbage in, garbage out" (GIGO) becomes acutely relevant. To address this challenge, our proposed framework centers around the delivery of reliable information to the machine learning model to enable precise predictions. Notably, this information is securely stored on the blockchain in an immutable format, rendering it impervious to external tampering. Furthermore, this approach safeguards against inadvertent erasure of students' academic records.

Beyond addressing the issue of fraudulent credentials, our research extends its reach to the career prospects of graduates. The next step involves providing students with guidance regarding their potential career paths, grounded in their academic performance.

Additionally, the proposed framework also holds significant benefits for employers. With our solution, recruiters need no longer question the authenticity of applicants' claims. The institution's blockchain infrastructure is readily accessible to reputable enterprises, enabling them to efficiently verify the credentials of fresh graduates. This verification process, integrated with blockchain technology, is poised to become an integral component of these enterprises' hiring procedures. Furthermore, students can seamlessly publish their reliable information on platforms like LinkedIn, which specialize in managing individuals' professional identities, establishing a comprehensive and permanent record for future reference.

This research underscores the critical need to address the challenges of fraudulent credentials, the lack of precise career guidance, and the verification of academic records in the education sector. By amalgamating blockchain and machine learning technologies, our proposed system, aptly named PRECOGNITO, strives to create a secure, trustworthy, and insightful ecosystem for students, educational institutions, and employers alike, fostering an environment of confidence and competence.

SIGNIFICANCE OF PROJECT

Degree fraud has emerged as a pervasive global challenge, exacerbated by the increasing demand for academic certificates in the competitive job market. This issue not only compromises the credibility of certifications issued by educational institutions but also introduces vulnerabilities in the verification process. [1] mentioned that countless degrees of fraud have occurred in the past. Traditional methods of degree verification rely on intermediaries, including certificate holders, educators, and university officials, leaving room for fraudulent activities such as counterfeit diplomas and forged degree certificates [7].

In response to these critical challenges, we introduce PRECOGNITO, a revolutionary system that harnesses the power of blockchain and machine learning technologies to redefine the landscape of education and employment verification. PRECOGNITO offers a multifaceted contribution to address the pressing issues within the domain:

Enhanced Data Security

PRECOGNITO leverages the robust security features of blockchain technology to combat the rampant issue of degree fraud. By immutably recording academic records on the blockchain, the system ensures that these records remain impervious to unauthorized alterations or deletions. This fortified data security forms the bedrock of academic credential credibility and authenticity.

Streamlined Verification Process

PRECOGNITO revolutionizes the degree verification process by eliminating the need for intermediaries. Employers gain direct access to the blockchain, facilitating swift and foolproof verification of prospective employees' academic qualifications. The system's ability to counter fraudulent degrees bolsters the integrity of hiring practices.

Empowering Informed Career Decisions

The integration of machine learning within PRECOGNITO empowers students to make informed and strategic career decisions. Utilizing predictive algorithms, the system provides valuable insights into future employment opportunities based

on individual academic performance. This feature equips students with the tools to chart career paths that align with their strengths and aspirations.

Fostering Collaborative Stakeholder Engagement

At its core, PRECOGNITO fosters a collaborative ecosystem where stakeholders, including students, educational institutions, and employers, actively engage in the verification process. This collaborative synergy ensures not only the accuracy and security of academic records but also the cultivation of trust among all parties involved.

In essence, PRECOGNITO represents a groundbreaking contribution to the realms of education and employment verification, offering innovative solutions to the pressing challenges faced. By addressing the vulnerabilities inherent in traditional verification methods and empowering students and employers alike, PRECOGNITO aspires to usher in a new era characterized by transparency, security, and informed decision-making in both educational and professional spheres. Through these transformative contributions, PRECOGNITO endeavors to reshape the landscape of education and employment verification, ultimately benefiting all stakeholders involved.

LITERATURE REVIEW

This section discusses the systems with the similarity with the system that we propose. This topic only touches on some samples of the many proposed improvements to the educational system and healthcare. Their varied and ever-expanding use cases are detailed, along with the technologies that enable them (such as blockchain and machine learning).

Degree Fraud

Degree fraud is an increasing problem in educational institutions. Mascot of the Diploma Mill Police Department, Chester Ludlow, a pug dog, was presented with an online Master of Business Administration degree from Rochville University in June 2009 for his "work and life experience." After sending $499 to the "university" in the United Arab Emirates by express mail, the dog received his diploma one week later (he could have graduated with honors for an additional $100) [8]. In another case, a North Carolina "doctor" is responsible for the death of an 8-year-old diabetic girl who was recommended to her mother to stop taking insulin. The "doctor," Laurence Perry, claimed to be an expert with degrees from a number of bogus online institutions. They were eventually traced back to fraudulent certificate vendors. After being found guilty of manslaughter and unlicensed medical practice, Perry received a 15-month prison term [9].

Platforms Utilizing the Blockchain

Student data validity and security are one of the main challenges that blockchain technology helped to address, which will significantly impact the future of the education area. For verification and safety, numerous strategies have been devised. The following are some concrete applications of Blockchain technology in education.

Secure Academic Certificate Authentication with Blockchain

The idea of flashing the system was introduced by [10] to prevent faking student degrees. The system incorporates cutting-edge cryptographic protocols like multi-signature authentication, a revocation mechanism based on the state of a Bitcoin (BTC) address, and trusted federated identification. The multi-signature method stands out among these protocols since it requires the signatures of most academic committee members each time a document is issued. Furthermore, because various devices and persons hold the private keys adopted by the multi-signature scheme, it improves the security of private key storage. BTC addresses are permanent and easily available, so the certificate revocation procedure that relies on them is more secure. To a significant extent, software like this has addressed the problems with the status quo, making the switch to a blockchain-based certificate more practical.

Figure 1 shows the overall system architecture of the proposed system, which consists of four components: The verification applications, the issuing application, the Blockchain, and a local Database.

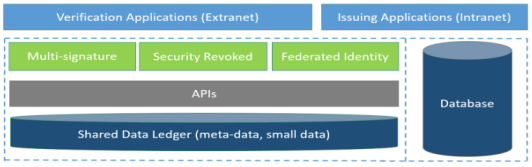

FIGURE 1. System Architecture Overview

An Electronic Certificate Infrastructure Based on the Blockchain

Due to the immutability of the blockchain, [11] have proposed a system with digital certificates that are stored and verified in the blockchain. With this strategy, important details are preserved, and a hiring manager cannot question an applicant's qualifications. Keeping track of one's education and development is essential for success in one's chosen field. Certificates help people achieve their academic goals. For this reason, recipients should keep their paper certificates safely stored and show them at every opportunity. Moreover, these documents should be kept in permanent, tamper-proof ledger books. An electronic certificate under this system is issued in accordance with the steps outlined below. Start by entering the digital copy of a paper certificate and its associated data into the database and then determine the hash value of the digital copy. The final step is to shop the hash value in the distributed ledger block. The system generates a QR code accompanying the certificate and an inquiry string code [3]. The demand unit will be able to verify the paper certificate by scanning it with a mobile device or performing a web search [12].

The interaction between the students, the institution, and the business is depicted in Figure 2.

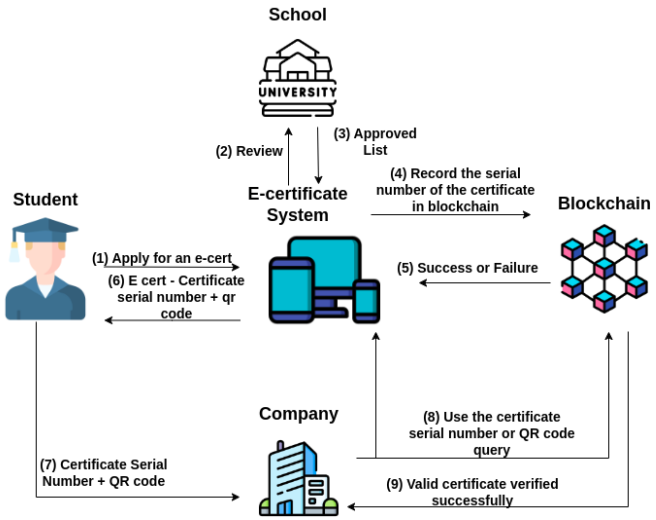

FIGURE 2. System Overview

Moreover, [13] offers a new way for individuals to manage their official transcripts and easily share them. The proposed approach uses blockchain technology, a distributed digital ledger of all cryptographic transactions. Blockchain technology's decentralized ledger design eliminates the requirement for a centralized authority to maintain or update data [7]. Han's proposed framework allows educational institutions to issue formal certificates as proof of completion or performance. It can also be utilized for further study beyond the four-year university level. This approach is a conduit between the university and its constituents (graduate students and outsiders). A consensus method is built into each node of the blockchain network; once a node has agreed that a transaction is valid, it will be permanently added to the blockchain (public ledger).Each client can connect to the distributed ledger and access the stored files independently. The person's private key decrypts all of the previous records.

Blockchain based healthcare record system

This article[14] delves into the transformative power of blockchain in healthcare, focusing on its impact on health record management, data security, and insurance billing. It presents innovative tools and an Access Control Policy Algorithm, highlighting blockchain's advantages over traditional Electronic Health Record (EHR) systems in terms of efficiency and security.The article introduces a blockchain-based Electronic Health Record sharing system, evaluating its performance and scalability using Hyperledger Fabric and Docker. Findings show that blockchain can enhance data collection, verification, and overall system security, heralding a potential revolution in healthcare systems.

This article[15] proposes blockchain technology for secure and efficient healthcare record management. It utilizes smart contracts to ensure data consistency and foster patient-physician coordination. The system prioritizes security, privacy, and real-time health data sharing.Operating on blockchain, the system enables digital access to patient records and secure medication history sharing. It employs blockchain and smart contracts for streamlined transactions and digital health record creation.The system securely stores patient data using SHA-256 hashing, with only data hashes on the blockchain. Patients access records through unique IDs. Authorized participants manage and share records, promoting decentralization.The system's evaluation covers resource utilization, contract response times, throughput, and latency. Its goal is secure, efficient, and transparent health record management, with future research focusing on system enhancement through a Modified Merkle Tree data structure.

The MedChain system[16] leverages blockchain technology for secure healthcare data management, effectively separating mutable and immutable data to enhance data integrity and security. Patients can securely share their data with healthcare providers, insurers, and research centers through cryptographic keys. MedChain records data events on the blockchain for tamper-proof and verifiable records, encrypts sensitive information for privacy, and employs a session-based sharing scheme. The system's components include super peers, edge peers, a blockchain service network, and a directory service for efficient data access.MedChain offers efficient, scalable data sharing with low storage and communication overhead. It excels in responsiveness and bandwidth consumption, demonstrating scalability in latency, responsiveness, and transaction rate. While the system requires manual patient approval for data access and manual data uploads to the blockchain, potential improvements include data transfer between healthcare organizations, automated sharing through authorized devices, and support for mobile and fog devices. Further research can refine these aspects, making MedChain an even more robust solution for healthcare data management.

Blockchain For Education

A cryptographically signed alternative to paper certificates is proposed in [6]. Institutions, students, and hiring managers can benefit from this technology's verification features. It promotes efficiency and security for certification authority by modernizing traditional processes, such as when a degree is confirmed on paper that lacks adequate proof and can be easily falsified. Certification authorities can boost efficiency while still protecting the integrity of their certificates due to this system's ability to issue and store certificates in a database and provide automatic certificate monitoring. Moreover, [6] advised protecting the certificate registry and using an open digital signature standard to verify global digital certificates.

Machine Learning-based System

A similar development in predicting students' performance and achievements have been observed in machine learning. The primary objective of machine learning is to forecast future events using historical data, also known as usable training data. This development in machine learning occurred at the same time as the development of predicting students' performance and achievements. Much study has been done to forecast students' academic progress and major accomplishments.

Foresight into Student Performance with Machine Learning Algorithms

Based on the Standard Interprofessional Data Mining Method, this work employs an approach that comprises six steps: Business understanding, Data Understanding, Data Preparation, Modelling, Evaluation of model, and Deployment [17]. The dataset used is obtained from the Scholar Management System *MASSAR*. Using the scikit-learn python package, the Multiple regression techniques generate the model in this system. The model has been divided into two parts: 1) a model for the second semester (Model S2) and 2) a model for the national examination (Model NE). ModelS2 predicts the second-semester results, while The National Examination model function forecasts the results of each student's national exam. The framework's impacts were evaluated in two stages. The initial stage of the process involves assessing standard measures such as mean absolute error (MAE), relative squared error, root means square error(RMSE), and relative absolute error. The subsequent phase involves using the results of an educational panel led by the head of the school. Moreover, [18] added that:"Robotic process automation (RPA) is a key aspect of machine learning in education since it can collect enormous amounts of data about students and tailor an experience to their needs" (p.3579).

Figure 3 represents the proposed framework for predicting students' performance which starts with collecting and preparing data from the SMS -Massar(School Management System "MASSAR").

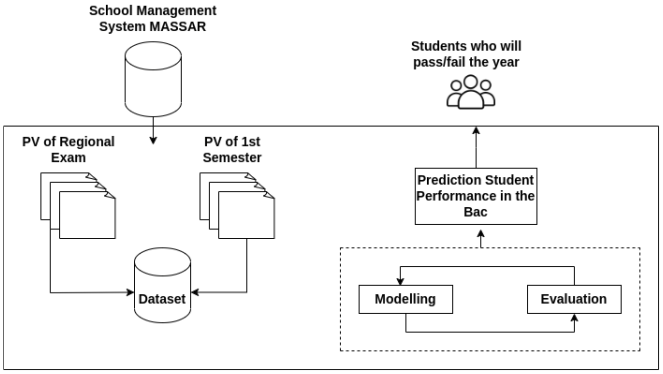

FIGURE 3. The Proposed Framework

The Application of ML to Instructional Design

Machine learning has a wide variety of intriguing and challenging uses in education. Given the breadth and depth of available information and the variety of student requirements, it stands to reason that a comprehensive ML system tailored to the specific requirements of a learning center will be popular in the twenty-first century. [19] was motivated to close the gap between reliable forecasts of student performance and traditional regression techniques. To predict student outcomes, this study used existing regression approaches. Many regression algorithms were examined to forecast student outcomes better and provide tutors with instructional help.

The suggested system will be rolled out in three stages in this project: the machine learning stage, wherein a model is proposed to forecast students' future employment; the Blockchain phase, when blockchain is built from the ground up to act as a database for storing student records in a somewhat secure manner, and the Integration phase, where the Machine Learning classifier is merged with the Blockchain to make them function together.

METHODOLOGY

The proposed PRECOGNITO system aims to steer the educational sector into the IR 4.0 direction of physical security and blockchain technology by combining both technologies' strengths (Blockchain and Machine Learning). To ensure that the proposed system meets the needs of parties involved in the education institutional management and accessibility to academic records (students, firm employers, colleges, and authorities), the researchers have considered their concerns about privacy, convenience, and availability.

FINDINGS

The PRECOGNITO system stands as a transformative and comprehensive solution that pioneers a new era in educational record management. Within this section, we delve into the primary discoveries and profound insights uncovered through our rigorous examination of the system's multifaceted components, intricate stakeholder dynamics, and cutting-edge technology implementations.

This comprehensive analysis underscores the system's pivotal role in reshaping conventional paradigms of educational record management. By merging blockchain and machine learning technologies, PRECOGNITO introduces a groundbreaking approach to address the longstanding challenges in this domain.

The PRECOGNITO Components

Integrating Blockchain and ML technologies and the roles and interactions of the many stakeholders are all depicted in Figure 4 as part of the overall system flow.

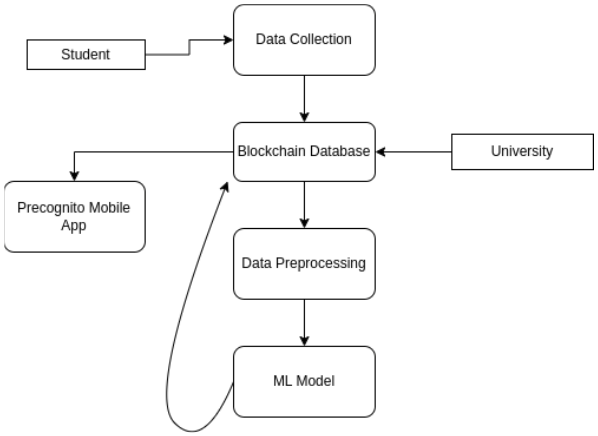

FIGURE 4. Flow of the proposed system

For clarity, the system is depicted hierarchically, with each level including a distinct group of stakeholders and the duties that pertain to them. Figure 5 presents a diagrammatic representation of the hierarchical structure of the proposed system.

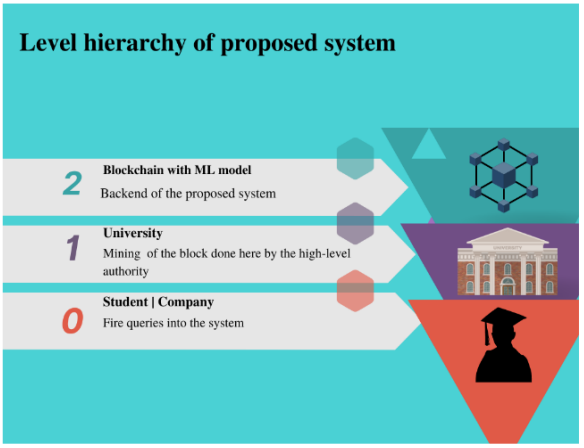

FIGURE 5. level hierarchy of the proposed system, the PRECOGNITO.

Student | Company

At the foundational level of the framework are students and companies, representing end-users who interact with the PRECOGNITO platform. Students can access their academic records, while companies can query the blockchain for student data, ensuring data accessibility and security.Students play a pivotal role in the PRECOGNITO framework as the primary beneficiaries. Their active participation involves the initial entry of their academic records into the blockchain. By providing accurate and reliable data, students ensure the integrity of their academic achievements. Furthermore, they can utilize the insights generated by the machine learning model to make informed decisions about their career paths. Companies(Employers) constitute another essential stakeholder group in the PRECOGNITO ecosystem. They benefit from the system's transparent and verified academic records, streamlining their hiring processes. Employers have direct access to the blockchain to verify the educational credentials of potential hires, ensuring the authenticity of applicants' qualifications.

University

The university functions as the administrative authority within the framework. It serves as an intermediary entity responsible for maintaining the integrity of the system. This role includes validating and verifying student data transactions within the blockchain. Additionally, the university oversees the system's smooth operation and resolves potential issues. Moreover,Educational institutions, represented by universities and colleges, serve as the custodians of student records. Their role is vital in maintaining data accuracy and security. Institutions are responsible for verifying and validating student data transactions on the blockchain. They also oversee the smooth operation of the system and address any technical challenges that may arise.

Blockchain with ML model

The highest level of the framework is the decentralized database that combines blockchain and machine learning. This critical component is responsible for processing user-initiated queries, safeguarding data integrity, and facilitating data

retrieval. Should this component fail, the entire system's functionality could be compromised. Thus, meticulous care is taken in its management and maintenance.

Data Collection

The first step to take in any machine learning project is the collection of data which will be used to train the models. In this project, the survey is used as the data collection instrument, and the population sample is students who have already graduated from IIUM sample dataset can be found at /dataset on GitHub repo [20].

Blockchain

Blockchain technology is implemented using python. Python is not a problem-specific language. Therefore, it may be used to create a wide variety of programs. The blockchain class will contain numerous methods, each of which will do a specific action, such as registering a new block, authenticating the chain, adding a new node, adding additional transactions, etc. It is also possible to use the secure hash algorithm 256 to determine a block's hash (SHA256).

Code for generating a proof of work, authenticating the chain, and adding a block to the chain is displayed in Figure 6. In this scenario, the transactions are the student records. The process accepts a Python dictionary as input and produces a newly constructed block that can be added to the existing chain.

```
def add_block(self, transactions):
    previous_hash = (self.chain[len(self.chain) - 1]).hash
    new_block = Block(transactions, previous_hash, str(datetime.datetime.now()))
    # calculate nonce
    proof = self.proof_of_work(new_block)
    # new_block.hash = proof
    self.chain.append(new_block)
    return proof, new_block
```

FIGURE 6. Code to add a new block to the chain

Data Preparation

The dataset[20] used in this work contains 15 independent features and 1 dependent feature, a categorical value. Three more categorical variables in the dataset must be converted to numeric values by factorization and categorical encoding to apply a specific Machine Learning technique to the dataset. The future job role of students is the key variable of interest in this dataset. Hence, the student's job is the target variable, dependent on the remaining variables.

The dataset is collected through a Google Form survey of current and former IIUM students. The survey also takes into account information found in a student's permanent records at the university. Therefore, many of the responses were not parallel. To fix the issues, such as unwanted spaces were replaced using regular expressions. Then, rather than removing rows containing duplicate values acting as unique values, those values were replaced with actual values. For example, the "loc" method of the pandas DataFrame is used to find rows in the "JOB ROLE" column where the value is "Network engineer." Once the rows are chosen, the value "Network engineer" in the "JOB ROLE" column is changed to "Network engineer." This is a case correction operation, where all instances of "Network engineer" in the specified column of the DataFrame "df" are changed to "Network Engineer."

MODEL TRAINING, SELECTION AND EVALUATION

In this section, we describe the process of model training and the rationale behind the selection of the Random Forest (RF) algorithm as our primary machine learning approach. We also provide an overview of the other models considered and their respective performance metrics.

Data Splitting

Before delving into the details of model training and selection, it's essential to clarify the division of our dataset into training and testing subsets. To ensure an unbiased evaluation of model performance, we randomly split our dataset into two portions: a training set and a testing set.

The training set comprised 70% of the total dataset, allowing our models to learn from a substantial portion of the data. This larger training set size contributed to robust model training and feature learning. The remaining 30% of the dataset was dedicated to the testing set. This separate testing set was utilized to assess the generalization capability of our models, measuring their performance on unseen data.

Model Selection

The choice of an appropriate machine learning algorithm plays a pivotal role in achieving accurate and reliable results. After careful consideration and a review of the literature, we opted for the Random Forest (RF) algorithm as our primary modeling approach. RF is a widely acknowledged ensemble learning technique known for its robustness and effectiveness in diverse domains [21]. The decision to use RF is supported by numerous studies in the literature that have demonstrated its superiority over other algorithms in classification tasks [21].

In particular, RF excels in handling high-dimensional data and is less prone to overfitting compared to single decision trees. It combines the predictions of multiple decision trees, each trained on a different subset of the data, resulting in an ensemble model that generalizes well to unseen data. Additionally, RF provides important insights into feature importance, which can aid in understanding the underlying factors contributing to the classification results [21]).

The model's performance metrics, including accuracy, precision, recall, and F1-score, showcase the effectiveness of RF in our dataset. With an accuracy of (0.936 – *Table 1.0)* and high precision and recall values, RF demonstrates its capability to accurately classify instances, which is essential in our application.

Model Evaluation

To evaluate the models, we employed a range of performance metrics, including accuracy, precision, recall, and the F1-score. These metrics provide a comprehensive view of each model's ability to classify instances accurately and deal with imbalanced data, which is often the case in classification problems.

Figure 7 illustrates the connection between the two variables as measured by the Pearson correlation coefficient [22]; Linear dependence measures the degree to which two variables are linked. Its value can be anywhere from -1 to 1, based on the following factors: A -1 correlation indicates a negative linear relationship when comparing two variables. If both values are 0, there is no linear relationship between them. If the correlation coefficient between two variables is 1, it indicates a perfect linear link between them. The distance that a correlation coefficient for a set of variables traverses away from zero indicates the intensity of the relationship between those variables. Furthermore, those who spend more time-solving coding problems on Code force (a platform to operate, organize, and discuss programming contests) are more likely to spend time participating in programming contests, as shown in the figure, which displays a correlation between "P_contest" and "Code_Force_PRACTICE" of 0.83.

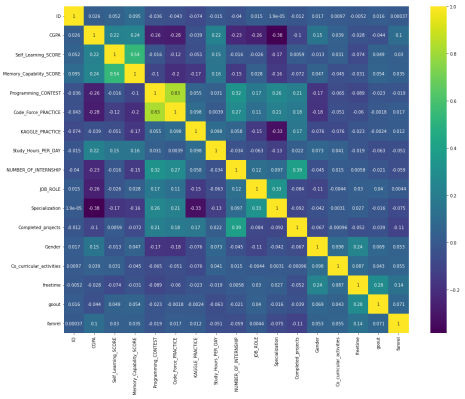

FIGURE 7. Correlation matrix of the dataset
TABLE 1. Summary of results

Model Name	Accuracy	Precision	Recall	F1-score
Random Forest	0.936	0.973	0.960	0.960
Stacking Classifier	0.852	0.731	0.776	0.842
Logistic	0.404	0.169	0.190	0.334
Gradient Boost	0.904	0.925	0.948	0.924

Figure 8 represents the ML model's top 10 most important features. "Specialization" feature has the highest importance in the ML model, indicating that a student's area of specialization significantly impacts the target variable "JOB_ROLE."

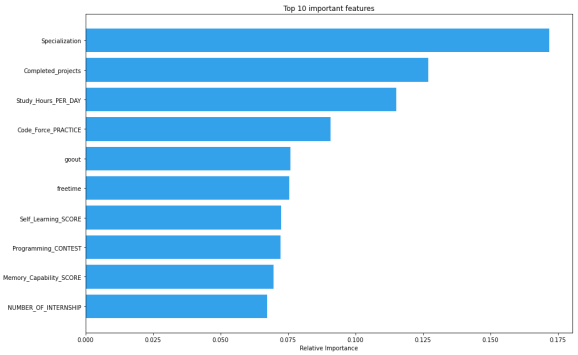

FIGURE 8. Top 10 most important features

Integration of Machine Learning and Blockchain

The fusion of Machine Learning (ML) and Blockchain technologies through a Flask-based Application Programming Interface (API) signifies a pivotal step toward creating innovative, secure, and efficient data-driven applications.

This integration bridges two transformative technologies: Blockchain, known for its data integrity and security, and ML, renowned for its data analysis and prediction capabilities. The Flask API serves as a vital conduit for seamless communication between these technologies, the key objectives include using ML to enhance data analysis within the Blockchain ecosystem. This integration enables real-time fraud detection, market trend prediction, and the execution of smart contracts powered by ML models.

In essence, this Flask-based API opens doors to a new era of intelligent, data-driven applications, leveraging the strengths of both ML and Blockchain while maintaining efficiency and security.

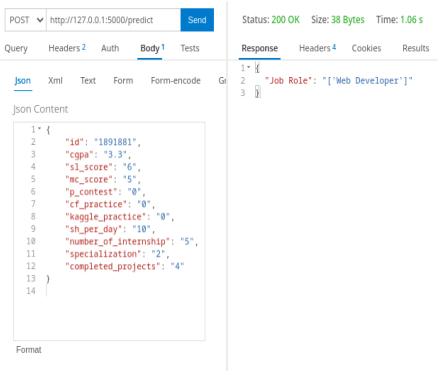

FIGURE 9. POST request sent from Thunder Client to the Flask server's predict route

The results of the API request discussed earlier can be seen in Figure 9. The /predict route receives a post request with 11 JSON characteristics of the 'id' '1891881'. 'Web Developer' is the JSON response received after posting to the Flask server API. This fusion of two technologies offers a fresh take on safety for traditionally implemented centralized database systems. Data from a fully authentic source, like Blockchain, is crucial for model training because of its verifiability and correctness.

Testing

The testing phase would begin after the implementation and involve all functional testing. The goal of the test is to check that the requirements have been met and that the system is functioning correctly. For blockchain-based components, these are the features tested:

- Functional testing: It is a comprehensive process that assesses the performance of the Blockchain's functional components.
- Application Programming Interface (API) testing examines how applications interact in the blockchain environment. It verifies that API requests and responses are appropriately structured and handled.
- Performance testing: detects performance bottlenecks, proposes ways to fine-tune the system, and determines whether the programmer is ready to go live.
- Node testing: To achieve smooth cooperation, all heterogeneous nodes on the network must be checked independently.

For Machine learning, these are some of the features we tested: A dataset of 1177 students was used to train the ML model. It has 11 independent features and 1 dependent feature. A Random Forest classifier has been used to prepare the ML model. Table 2 illustrates the result from the ML model testing.

TABLE 2. ML model testing

Input	Expected result	Status
Student Data	Predicting Jobs	89% accuracy

CONCLUSION

In summary, our project, PRECOGNITO, represents a pivotal milestone in educational technology, offering a comprehensive system that harmonizes blockchain and machine learning to address critical challenges in the education sector.

Our primary contribution lies in the holistic PRECOGNITO system, where the symbiosis of two remarkable technologies sets the stage for a paradigm shift in educational management. By integrating blockchain as a secure foundation and leveraging machine learning for predictive analytics, we provide a multifaceted solution.Blockchain enhances academic record security, fortifying the credibility of certifications while simplifying the verification process. Simultaneously, machine learning empowers students with predictive insights into future career opportunities based on academic performance.

Central to our vision is the creation of a decentralized student data management system, enabled by blockchain technology, offering efficiency and transparency for all stakeholders.

In essence, PRECOGNITO is a testament to the profound synergy between blockchain and machine learning, offering tangible benefits to students, educational institutions, and employers alike. It encapsulates a transformative step towards a more secure, data-driven, and future-oriented educational landscape.

Our research contributes significantly to the ongoing discourse on the advancement of educational technologies. As we look ahead, we envision PRECOGNITO as an indispensable tool in shaping the future of education, instilling trust, and empowering students to make well-informed decisions about their career paths.

FUTURE WORK

The model can be enhanced by employing new algorithms such as logistic regression, support vector machine, extreme gradient boosting, decision tree, etc., leading to better results or accuracy. A high-fidelity system prototype will be built and sent out to the stakeholders for evaluation. An overview of the system before the main development phase (Mobile App) is done in the future.

REFERENCES

1. Chen, G., Xu, B., Lu, M., & Chen, N.-S. (2018). Exploring blockchain technology and its potential applications for education. *Smart Learning Environments*, 5:1.
2. Schwardmann, U., Boehme, C., Heras, D. B., Cardellini, V., Jeannot, E., Salis, A., ... & Scott, S. L. (Eds.). (2020). Euro-Par 2019: Parallel Processing Workshops: Euro-Par 2019 International Workshops, Göttingen, Germany, August 26–30, 2019, Revised Selected Papers (Vol. 11997). Springer Nature.
3. Shah, D., Patel, D. H., Adesara, D. J., & Hingu, P. (2021). Exploiting the Capabilities of Blockchain and Machine Learning in Education. *Augmented Human Research*, 6(1).
4. A-Primer-on-Machine-Learning.-(n.d.).A-Primer-on-Machine-Learning.-Https://ce.uci.edu/about/magazine/articles/fa17_machineLearning.aspx
5. Abroyan, & Hakobyan. (2016). A Review of the Usage of Machine Learning in Real-time Systems. *Test.journals.sci.am*.
6. Gräther, W., Kolvenbach, S., Schütte, J., Ruland, R., Torres, C., & Wenland, F. (2018). Blockchain for Education: Lifelong Learning Passport. *European Society for Socially Embedded Technologies (EUSSET)*.
7. Lu, Y. (2018). Blockchain: A survey on functions, applications, and open issues. *Journal of Industrial Integration and Management*, *3*(04), 1850015.
8. Murrow, J. (2020, January 21). Pug Dog Earns Online MBA in Rochville University Scam. GetEducated | Review, Rate,Rank-&-Compare-Online-Colleges-&-Degrees|GetEducated.https://www.geteducated.com/diploma-mill-police/dog-earns-online-mba/
9. Gilbert, K. (2020, June 2). Top 10 Instant Degree Scams | GetEducated. GetEducated | Review, Rate, Rank & Compare-Online-Colleges-&-Degrees|GetEducated.https://www.geteducated.com/life-experience-college-degree/290-online-college-diploma-mill-cases/
10. Li, R., & Wu, Y. (2018). Blockchain-based Academic Certificate Authentication System Overview. *IT Innov. Centre, Univ. Birmingham*, 8.
11. Gopal, N., & Prakash, V. V. (2018). Survey on Blockchain-Based Digital Certificate System. *International Research Journal of Engineering and Technology (IRJET)*, 5.
12. Cheng, J. C., Lee, N. Y., Chi, C., & Chen, Y. H. (2018, April). Blockchain and smart contract for digital certificate. In *2018 IEEE international conference on applied system invention (ICASI)* (pp. 1046-1051). IEEE.
13. Han, M., Li, Z., He, J., Wu, D., Xie, Y., & Baba, A. (2018, September). A novel blockchain-based education records verification solution. In *Proceedings of the 19th annual SIG conference on information technology education* (pp. 178-183).
14. Bhattacharya, P., Tanwar, S., Bodkhe, U., Tyagi, S., & Kumar, N. (2019). Bindaas: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications. *IEEE transactions on network science and engineering*, *8*(2), 1242-1255.
15. Chelladurai, U., & Pandian, S. (2022). A novel blockchain based electronic health record automation system for healthcare. *Journal of Ambient Intelligence and Humanized Computing*, 1-11.
16. Bhattacharya, P., Tanwar, S., Bodkhe, U., Tyagi, S., & Kumar, N. (2019). Bindaas: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications. *IEEE transactions on network science and engineering*, *8*(2), 1242-1255.
17. Siddiqui, A. A. (2020). The Epidemiology of COVID-19 Novel Corona Virus in Incidence and the Distribution of the Disease across the World.
18. Qazdar, A., Er-Raha, B., Cherkaoui, C., & Mammass, D. (2019). A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Education and Information Technologies*, 3577–3589.
19. Kotsiantis, S. B. (2012). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artificial Intelligence Review*, 331-344.
20. Murtaj, S. M. T., & Yusif, A. A.. (2022). PreCogNito GitHub Repository. [GitHub Repository]. https://github.com/tahsinmurtaj16/PreCogNito
21. Breiman, L. (2001). Random forests. *Machine learning*, *45*, 5-32.
22. Benesty, J., Chen, J., Huang, Y., & Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing* (pp. 1-4). Springer, Berlin, Heidelberg.