SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Precognito: The Emergence of Blockchain & Machine Learning-Based for Student Record Authentication System

Sheikh Mohammad Tahsin Murtaj[1], Mohammed Mohammed[1], Abubakar Abubakar Yusif[1], Norzariyah Yahya[1,*], Rawad Abdulghafor[2]

[1] Department of Computer Science, Kulliyah of Information and Communication Technology, International Islamic University Malaysia, 53100 Jln Gombak, Malaysia
[2] Faculty of Computer Studies (FCS), Arab Open University–Oman, Muscat P.O. Box 1596, Oman

ARTICLE INFO

Article history:
Received 10 February 2024
Received in revised form 19 September 2024
Accepted 10 October 2024
Available online 15 January 2025

Keywords:
Precognito; blockchain; machine learning

ABSTRACT

Education is progressing more rapidly than ever due to technological advancements. Technological advancements have led to two problems: counterfeit transcripts and degrees, primarily caused by data security vulnerabilities. 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. The technology can prevent additional fraud and uncertainty in student achievements by storing student data on the Blockchain and leveraging machine learning techniques for precise analysis. It can enable accurate prediction of future employment opportunities for graduates. Machine learning algorithms are used to train and make accurate predictions, and the requisite data are retrieved from a Blockchain ledger. PRECOGNITO will equip the institution with a decentralised and immutable alumni database that contains verified and transparent academic records. Additionally, this system provides employers with a means to verify the validity 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.

1. Introduction

In today's automated society, the education industry leads in embracing innovative technologies, particularly blockchain and machine learning, which are seen as transformative forces with uses across multiple sectors. Blockchain and machine learning, among the most transformative technologies of the 21st century, are still not fully leveraged in education. This vital sector faces degree fraud, cybersecurity threats, and the disparity between academic credentials and job readiness skills.

* *Corresponding author.*
E-mail address: norzariyah@iium.edu.my

https://doi.org/10.37934/araset.55.1.125140

To tackle these issues, we proposed and developed a PRECOGNITO system that integrates blockchain and machine learning to create a secure, transparent platform for storing, managing, and analysing student records. We hypothesise that this platform can alleviate degree fraud, secure student data, offer precise performance assessments, and connect students with suitable employers. By utilising a mixed-methods approach, we seek to assess the feasibility and effectiveness of our proposed solution.

2. Related Works

Blockchain technology in education has garnered interest, particularly in creating a secure global assessment platform to store and manage degree information, as noted by Chen *et al.*, [1]. Schwardmann *et al., [2]* stated that blockchain project inherent data immutability makes it ideal for safeguarding academic credentials with universities actively exploring blockchain to maintain students' records and facilitate sharing potential candidates' details with potential employers, as Shah *et al., [3]* suggested.

Machine learning (ML) is a distinct area within computer science that utilises recursive learning processes to glean insights from data [4]. However, Machine learning encounters difficulties due to unreliable or contaminated data, as stated by Abroyan and Hakobyan [5], highlighting the need for our suggested framework. This framework combines blockchain's secure storage capabilities with machine learning to ensure accurate predictions. This section delves into systems similar to ours, touching on samples of proposed educational and healthcare improvements. They provide a comprehensive overview of their diverse and continually growing use cases and the enabling technologies, including blockchain and machine learning.

Educational institutions are increasingly dealing with degree fraud. For example, Chester Ludlow, a pug dog, was awarded an online MBA, as noted by GetEducated [6]. Additionally, a "doctor" in North Carolina faced serious consequences for holding degrees from fake online institutions, as reported by Collier [7]. Moreover, a cryptographically signed alternative to paper certificates is proposed by Gräther *et al.* [8], addressing the need to protect the certificate registry and use an open digital signature standard to verify global digital certificates.

Blockchain technology enhances the validity and security of student data. It facilitates to secure academic certificate authentication, as demonstrated by Li and Wu [9], along with an electronic certificate infrastructure suggested by Gopal and Prakash [10], and systems for managing healthcare records. Cheng *et al*. [11] introduce blockchain technology for secure and efficient healthcare record management, utilising smart contracts to ensure data consistency and foster patient-physician coordination. The system prioritises security, privacy, and real-time health data sharing. Operating on a blockchain, the system enables digital access to patient records and secure medication history sharing. Similarly, Han *et al.* [12] offered a new way for individuals to manage their official transcripts and easily share them. Their suggested method employs blockchain technology, a decentralised digital ledger for all cryptographic transactions, to ensure secure and efficient record management beyond the university level.

Figure 1 shows the overall system architecture of the proposed system by Li and Wu [9], which consists of four components: the verification applications, the issuing application, the blockchain, and a local database.

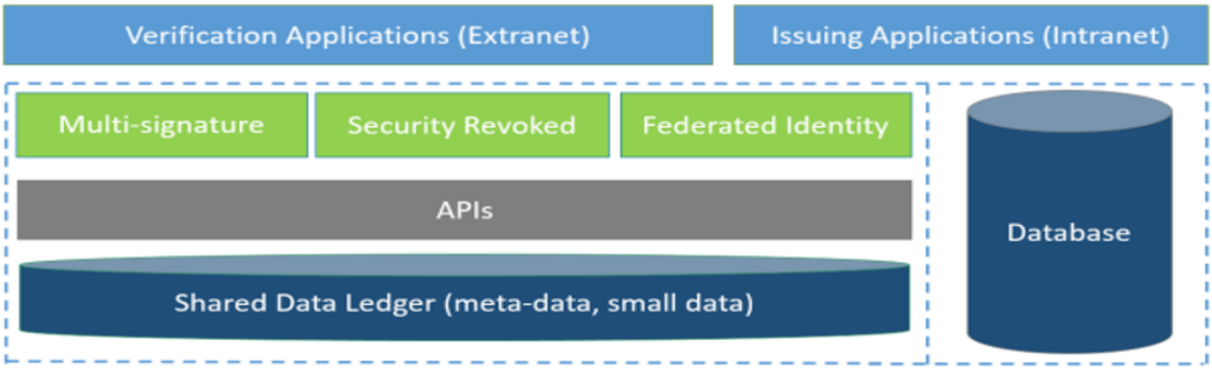

Fig. 1. System architecture overview [9]

Figure 2 illustrates the working process of blockchain technology in producing e-certificates for students in an educational institution. The interaction between the students, the institution, and the company that manages the e-certificates is the proposed system by Cheng *et al.,* [11]. The process started with the students applying for an e-certificate, which involved the educational institution reviewing and approving the list of graduates. A decentralised blockchain database records the e-certificates for students who have requested e-certificates. Upon completing the procedure detailed in Figure 2, students will receive a copy of their e-certificates.

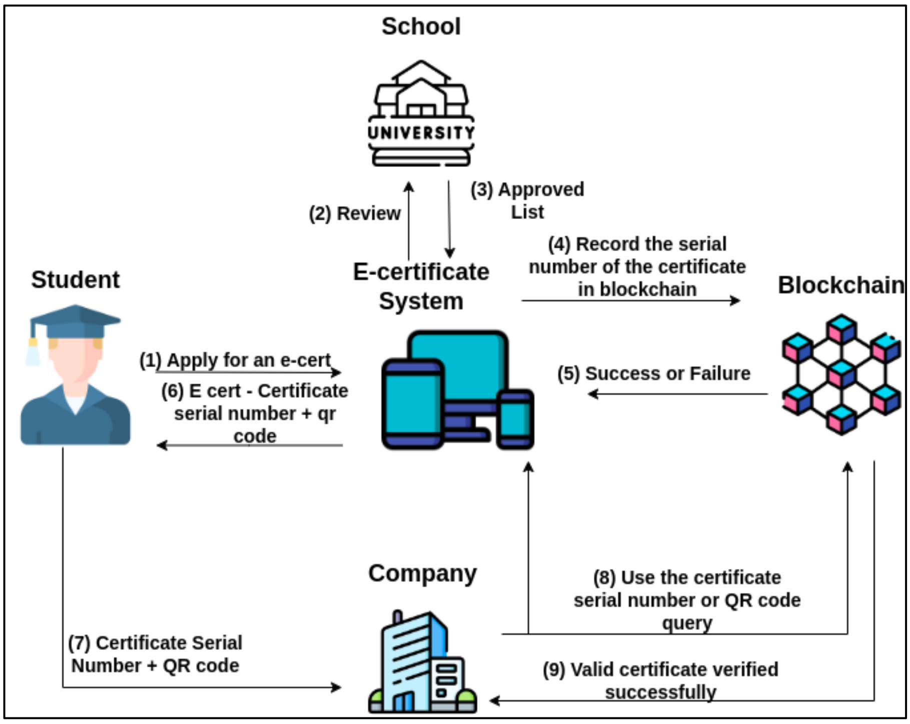

Fig. 2. System overview [11]

Additionally, Bhattacharya *et al.,* [13] explored the transformative power of blockchain in healthcare, focusing on its impact on health record management, data security, and insurance billing. It introduces innovative tools and an Access Control Policy Algorithm, emphasising the advantages of blockchain over traditional Electronic Health Record (EHR) systems in terms of efficiency and security. The article introduces a blockchain-based Electronic Health Record sharing system and evaluates 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. Chelladurai and Pandian [14] proposed blockchain technology for secure and efficient healthcare record management. It utilises smart contracts to ensure data consistency and foster patient-physician coordination. The system prioritises security, privacy, and real-time health data sharing. Another example of using blockchain technology for healthcare data management is MedChain proposed by Shen *et al.,* [15]. This system separates mutable and immutable data to improve data integrity and security. MedChain allows patients to share their data securely with various stakeholders using cryptographic keys.

Machine learning predicts student performance by utilising robotic process automation (RPA), as stated by Qazdar *et al.,* [16] and aids instructional design suggested by Kotsiantis [17], emphasising the need for comprehensive ML systems tailored to learning centres in the twenty-first century. Moreover, Siddiqui [18] employs machine learning algorithms to provide foresight into student performance, utilising Scholar Management System MASSAR data. 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").

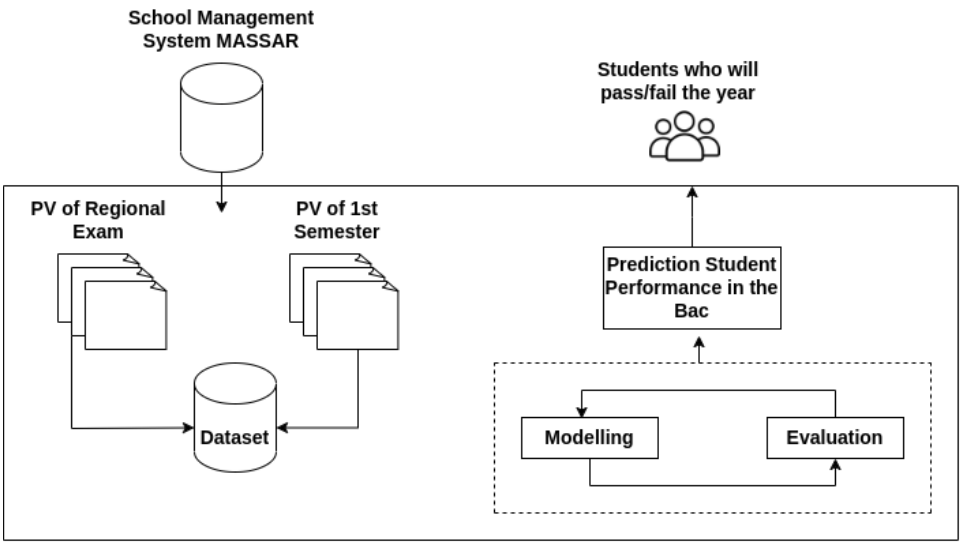

Fig. 3. The proposed framework of MASSAR [18]

While existing literature explores solutions in blockchain and machine learning individually, a clear gap exists for a comprehensive system that integrates both technologies.

Diverse educational challenges, such as fake degrees, credential forgery, and the absence of a standardised, secure, and decentralised database, highlight the necessity for a reliable information delivery system. Leveraging the insights from Gräther *et al.,* [8], our proposed solution aims to address these challenges comprehensively. Our project, PRECOGNITO, addresses the global challenge of degree fraud by leveraging blockchain and machine learning, building on the foundation of Lu [19] in enhanced data security. The significance lies in improving data security, streamlining verification, empowering students with predictive insights, and fostering collaborative stakeholder engagement. PRECOGNITO represents a transformative solution that reshapes educational record management and contributes to transparency, security, and informed decision-making in education and employment verification.

3. Methodology

The PRECOGNITO system is designed to guide the education sector towards Industry Revolution 4.0 (IR 4.0) by leveraging the strengths of physical security and blockchain technology, specifically integrating Blockchain with 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. The next two sections explore the elements of the PRECOGNITO, shown in Figure 4, and the PRECOGNITO process, illustrated in Figure 5.

3.1 The Precognito Components

The system is represented hierarchically for enhanced clarity, with each level comprising a specific group of stakeholders and their corresponding responsibilities. Figure 4 presents a diagrammatic representation of the hierarchical structure of the proposed system.

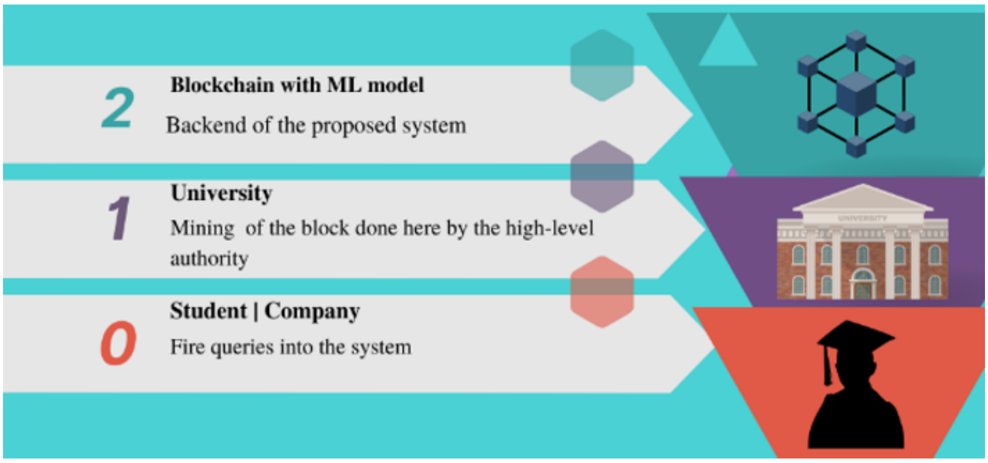

Fig. 4. Level hierarchy of the proposed system, the PRECOGNITO

3.1.1 Student or company

At the foundational level of the framework are students and companies representing end-users engaging 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 entering their academic records into the blockchain. Students uphold the integrity of their academic achievements by providing dependable and precise data. Furthermore, they can utilise 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 gain advantages from the system’s transparent and validated academic records, simplifying their hiring procedures. Employers are able to access the blockchain directly to authenticate the educational credentials of potential candidates, confirming the validity of their qualifications.

3.1.2 University

The university functions as the administrative authority within the framework. It serves as an intermediary entity responsible for maintaining the system's integrity. This role includes providing the data for the blockchain and verifying student data before it is stored in 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 manage the seamless functioning of the system and tackle any technical issues that come up.

3.1.3 Blockchain with ML model

The highest level of the framework is the decentralised database that combines blockchain and machine learning. This critical component is responsible for processing user-initiated queries, safeguarding data integrity, and facilitating data retrieval. If this component fails, it could jeopardise the entire system. Therefore, careful attention is prioritised in its management and upkeep.

Figure 5 illustrates the integration of Blockchain and ML technologies and the roles and interactions of various stakeholders within the overall system flow. The process flow of this project commenced with collecting data on students' transcripts and details of degrees awarded. Subsequently, the gathered data will be stored in a decentralised blockchain database, which utilises distributed ledger technology that incorporates smart contracts to retain students' transcript information and the degrees conferred upon them. The details of the process are illustrated in Figure 5.

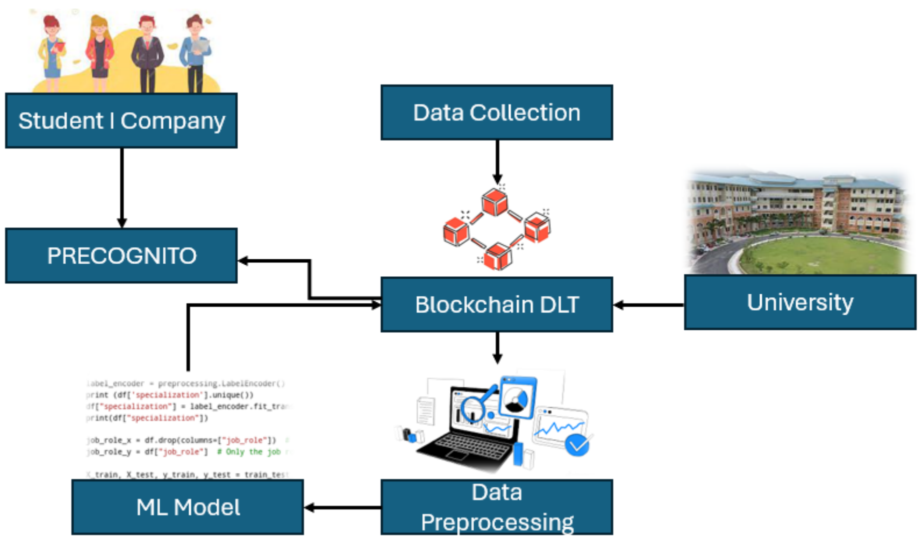

Fig. 5. Flow of the PRECOGNITO proposed system

3.2 Data Collection

The first step in any machine learning project is to collect data that 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 the International Islamic University Malaysia (IIUM). The sample dataset can be found at /dataset on the GitHub repo Murtaj and Yusif [24].

3.3 Blockchain

In this project, we use Python as the platform to create a smart contract to store and authenticate the data of students and graduates. Two key data sets include students' course grades and the degrees they have earned graduates. In this blockchain module of the project, all data is obtained by distributing a survey for students to record their grades for courses taken and degrees for those who have graduated. It should be noted that PRECOGNITO is an independent initiative aimed at harnessing blockchain and machine learning technologies for the secure storage and management of degree information, helping to prevent degree forgery.

Additionally, Python is favoured because its language is not limited to specific problems, enabling the development of a diverse range of programs. Besides the main objective of the smart contract, which is to store and authenticate the students' and graduates' data, the blockchain modules in PRECOGNITO consist of classes that 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. Using the secure hash algorithm 256 is also possible 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.

```python
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
```

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

The type of blockchain used in this project is a private permissioned blockchain based on the Hyperledger Fabric framework. This blockchain suits the projects' goal of securely, efficiently and reliably processing and storing student records [20]. It also enables faster and cheaper transactions, as it does not need a resource-intensive consensus algorithm like proof of work, according to Seth [21]. Moreover, it supports smart contracts, which automate the systems' logic and rules records [20]. This blockchain has the advantage over other blockchains in security, efficiency, and scalability, as stated by Seth [21]. It prevents unauthorised access, data tampering, and attacks, as suggested by MacDonald [22]. It reduces transaction time, cost, and complexity. It handles more transactions and data without affecting performance and quality [23].

3.4 Data Preparation and Preprocessing

The dataset utilised in this study by Murtaj and Yusif [24] includes 15 independent features and 1 dependent feature, a categorical value. The dataset requires the conversion of three additional categorical variables into numeric values through factorisation and categorical encoding to implement a particular machine-learning technique. 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 from current and former IIUM students through a Google Form survey. The survey also considers information from a student's permanent records at the university. As a result, many responses were not consistent. To address this, unwanted spaces were removed using regular expressions. Instead of deleting rows with duplicate values that were supposed to be unique, those values were replaced with the correct ones. For instance, the pandas DataFrame's "loc" method is utilised to locate rows in the "JOB ROLE" column that contain the value "Network engineer." After selecting these rows, the value "Network engineer" in the "JOB ROLE" column is updated 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."

3.5 Machine Learning: Model Training, Selection and Evaluation

This section describes the model training process and the rationale behind selecting 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.

3.5.1 Data splitting

Before delving into the details of model training and selection, it is essential to clarify the division of our dataset into training and testing subsets. To ensure an unbiased model performance evaluation, we randomly split our dataset into two portions: a training set and a testing set. The training set comprised 70% of the dataset, allowing our models to learn from a substantial portion. 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. The distinct testing set evaluated how well our models generalise, assessing their performance on unseen data.

3.5.2 Model selection

Choosing an appropriate machine learning algorithm is pivotal 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 modelling approach. RF is a widely acknowledged ensemble learning technique known for its robustness and effectiveness in diverse domains, as stated by Breiman [25]. The decision to use RF is supported by numerous studies demonstrating its superiority over other algorithms in classification tasks, as presented in Breiman [25].

RF is particularly adept at managing high-dimensional data and is less susceptible to overfitting than individual 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 generalises 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, as presented in Breiman [25]. the model's performance metrics, including accuracy, precision, recall, and F1-score, demonstrate the effectiveness of RF in our dataset. Achieving an accuracy of 0.936 (as shown in Table 1) alongside remarkable precision and recall values, RF showcases its proficiency in classifying instances, which is essential for our application.

3.5.3 Model evaluation

We employed a range of performance metrics to evaluate the models, 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 described by Benesty *et al.,* [26]. Linear dependence gauges how closely two variables are related. The correlation value ranges from -1 to 1, influenced by these factors: A correlation of -1 signifies a negative linear relationship between the two variables. Conversely, if both values are 0, no linear relationship exists between them. If the correlation coefficient between two variables is 1, it indicates a perfect linear link between them. The correlation coefficient's distance from zero reflects the strength of the relationship among the variables. Additionally, individuals who dedicate more time to solving coding challenges on Codeforces—a platform designed for organising and discussing programming contests—tend to engage more in these contests. This is evidenced by the figure showing a correlation of 0.83 between "P_contest" and "Code_Force_PRACTICE."

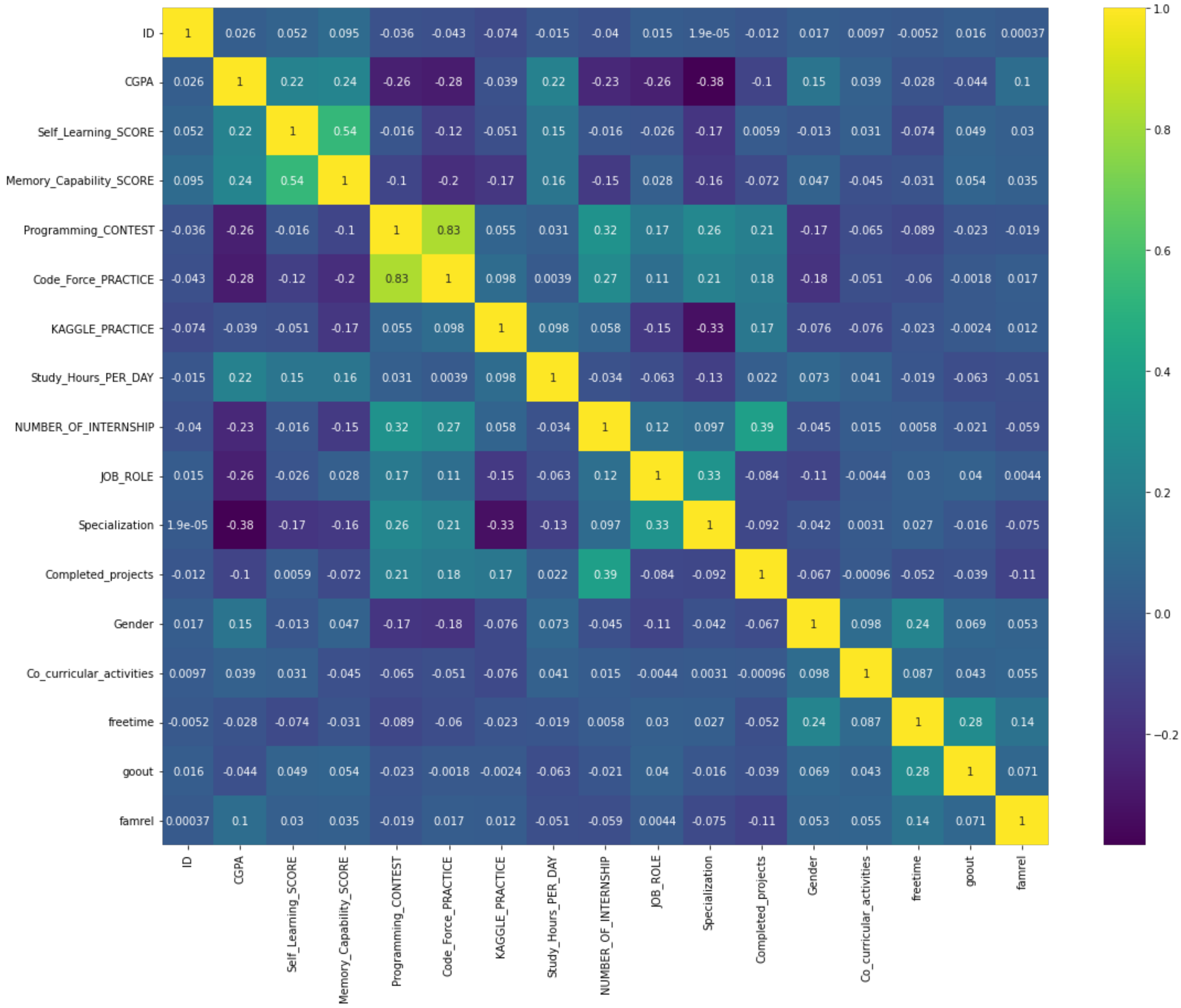

Fig. 7. Correlation matrix of the dataset

Figure 8 represents the ML model's top 10 most important features. The "specialisation" feature is most important in the ML model, indicating that a student’s area of specialisation significantly impacts the target variable "JOB_ROLE."

3.6 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 is 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.

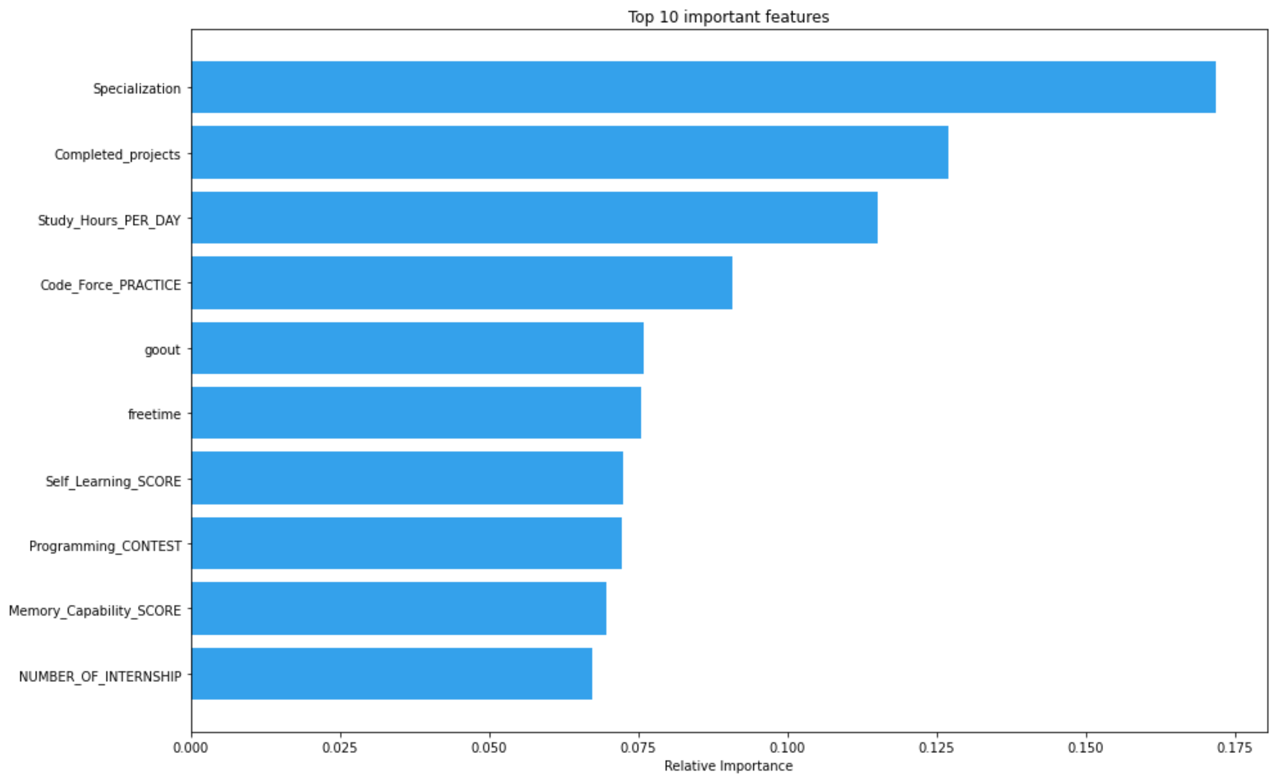

Fig. 8. Top 10 most important features

3.7 Testing

The testing phase will commence post-implementation and encompass all functional tests. Its purpose is to verify that the requirements are fulfilled and that the system operates as intended. For blockchain-based components, the features tested are:

i. Functional testing: It is a comprehensive process that assesses the performance of the Blockchain’s functional components.

ii. 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.
iii. Performance testing detects performance bottlenecks, proposes ways to fine-tune the system, and determines whether the programmer is ready to go live.
iv. Node testing is performed to ensure smooth cooperation; therefore, all heterogeneous nodes on the network must be assessed independently.

4. Results and Discussion

The PRECOGNITO system is a transformative and comprehensive solution that emerges two technologies, blockchain and machine learning, in a new era in educational record management. This section explores the key findings and deep insights revealed through our thorough investigation of the system's complex components, intricate stakeholder relationships, and advanced 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 blockchain is a component of PRECOGNITO that runs the smart contract to store students' and graduates' data. Blockchain ensures that the data is immutable and decentralised, and the distributed ledger technology allows the transactions to be transparent. However, with certain restrictions in access due to the confidentially of the data stored in the blockchain module, in PRECOGNITO, the modules are being developed as permissioned.

A robust validation process was employed to ensure the reliability and accuracy of the machine learning (ML) model. We utilised a k-fold cross-validation technique, specifically a 5-fold validation, to assess the model's performance. This method involves dividing the dataset into five subsets, training the model on four subsets and validating the remaining one in each iteration. The process was repeated five times, and the average performance metrics were computed. Applying k-fold cross-validation enhances the model's robustness by providing a more comprehensive evaluation across different subsets of the data, minimising the risk of overfitting.

After performing a comparative analysis with other models, such as Stacking Classifier, Logistic Regression, and Gradient Boost, the Random Forest classifier was chosen for our machine learning model. The rationale behind selecting the Random Forest classifier lies in its ability to handle complex relationships in the data, mitigate overfitting, and deliver high accuracy across diverse datasets. This careful consideration of alternative models supports our decision to employ Random Forest as the most suitable choice among the models evaluated. The comparative results demonstrated that the Random Forest classifier outperformed the alternative models' accuracy. This superiority can be attributed to its ensemble learning approach, which aggregates predictions from multiple decision trees, resulting in a more robust and accurate predictive model.

Table 1 summarises our machine-learning models' accuracy, precision, recall, and F1 score before incorporating cross-validation techniques.

Table 1
Accuracy, precision, recall, F1-score of machine learning models

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

We applied Stratified K-Fold Cross-Validation to assess the models' robustness. The results demonstrated a mean accuracy of 0.975 with a low standard deviation of 0.011, indicating consistent and reliable performance across varied data subsets.

Comparing model performance metrics before and after cross-validation revealed a marginal increase in mean accuracy and a reduction in standard deviation. This enhancement signifies improved stability and reliability of our models.

Figure 9 showcases the results of an API request, exemplifying the system's functionality. The /predict route processes a post request, yielding a JSON response predicting 'Web Developer' for the 'id' '1891881'. This amalgamation of blockchain and machine learning technologies establishes a secure and innovative alternative to centralised database systems, emphasising the significance of authentic data for model training.

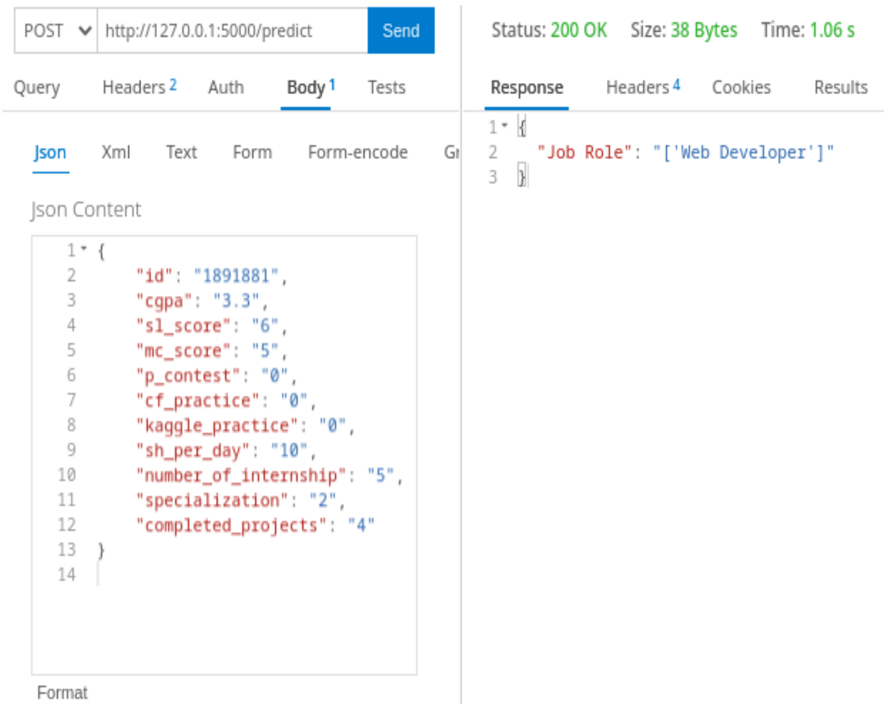

Fig. 9. POST request sent from thunder client to the flask server's predict route

Within the ML testing domain, our analysis utilised a dataset of 1177 students, employing 11 independent features and 1 dependent feature. The Random Forest classifier demonstrated notable accuracy, as detailed in Table 2.

Table 2
ML model testing

Input	Expected result	Status
Student Data	Predicting Jobs	93% accuracy

Table 3 presents the results of our performance testing scenarios, which aimed to evaluate the scalability and efficiency of our system under different workloads and database sizes. We used various parameters, tools, and metrics to simulate and measure the system's behaviour in realistic situations. The table shows that our system can handle low to high user activity, sudden load bursts, peak usage, and extended duration without significant response or query time degradation. The table also demonstrates that our system can process large volumes of data efficiently, as the query time does not vary much with the database size. These findings suggest that our system is scalable and efficient, meeting users' expectations.

Table 3
Scalability and performance analysis

Scenario	Objective	Parameters	Tools	Metrics
1. Low Workload	Evaluate system behaviour under minimal user activity.	Simulated a low number of concurrent users (e.g., 10).	JMeter	• Average Response Time: 20 ms • Minimum Response Time: 15 ms • Maximum Response Time: 25 ms
2. Burst Load	Assess how the system handles sudden spikes in user activity.	Rapidly increased the number of concurrent users (e.g., from 10 to 1000).	LoadRunner	• Average Response Time: 150 ms • Minimum Response Time: 120 ms • Maximum Response Time: 180 ms
3. Peak Usage	Mimic scenarios where the system experiences maximum usage.	Set workload patterns to simulate peak hours (e.g., 1000 users for 1 hour).	Gatling.	• Average Response Time: 180 ms • Minimum Response Time: 160 ms • Maximum Response Time: 200 ms
4. Extended Duration	Measure the system's performance over an extended period.	Run tests for an extended duration (e.g., 24 hours) and monitor for resource exhaustion or degradation.	JMeter	• Average Response Time: 220 ms • Minimum Response Time: 200 ms • Maximum Response Time: 240 ms
5. Database Size Impact	Evaluate the impact of different database sizes on system performance.	Used datasets of varying sizes (e.g., 10 MB, 100 MB, 1 GB, 10 GB, etc.).	Gatling	• Average Response Time: 30 ms • Minimum Query Time: 25 ms • Maximum Query Time: 35 ms

Incorporating these insights into our academic discourse enriches our understanding of the validation process, model performance, and the system's scalability, contributing to the broader discourse on educational record management systems.

5. Conclusion

PRECOGNITO offers a comprehensive system that emerges blockchain and machine learning to address critical challenges in the education sector. The primary contribution lies in the holistic PRECOGNITO system, where the emergence 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, PRECOGNITO provides a multifaceted solution. Blockchain enhances academic record security, fortifying the credibility of certifications while simplifying the verification process. At the same time, machine learning provides students with predictive insights regarding future career opportunities tied to their academic performance.

At the heart of our vision is developing a decentralised student data management system powered by blockchain technology. This system aims to provide efficiency and transparency for all stakeholders, including students, academic institutions, employers, and industries, ensuring access to accurate graduate data. PRECOGNITO showcases the deep connection between blockchain and machine learning, providing real advantages to students, educational organisations, and employers. It encapsulates a transformative step towards a more secure, data-driven, and future-oriented educational landscape.

The current work contributes to the ongoing discourse on the advancement of educational technologies. Looking ahead, PRECOGNITO is foreseen as an indispensable tool in shaping the future of education, with trust being instilled and students being empowered to make well-informed decisions about their career paths.

Despite the project's limitations and challenges, improvements to the current model can be achieved by employing new algorithms such as SVM, neural networks, and ensemble methods (AdaBoost), which may lead to better results or accuracy. The system requires further testing and evaluation from stakeholders, including students, educators, and employers, to ensure its usability and effectiveness. A high-fidelity system prototype is planned for development and distribution to stakeholders for feedback and validation. Additionally, a mobile app version of the system is intended to enhance user convenience and accessibility. These are potential future enhancements and directions for our system development.

Acknowledgments
This research was not funded by any grant.

References
[1] Chen, Guang, Bing Xu, Manli Lu, and Nian-Shing Chen. "Exploring blockchain technology and its potential applications for education." *Smart Learning Environments* 5, no. 1 (2018): 1-10. https://doi.org/10.1186/s40561-017-0050-x
[2] Schwardmann, Ulrich, Christian Boehme, Dora B. Heras, Valeria Cardellini, Emmanuel Jeannot, Antonio Salis, Claudio Schifanella et al. "Correction to: Euro-Par 2019: Parallel Processing Workshops." In *Euro-Par 2019: Parallel Processing Workshops: Euro-Par 2019 International Workshops, Göttingen, Germany, August 26–30, 2019, Revised Selected Papers 25*, pp. C1-C1. Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-48340-1_64
[3] Shah, Dhruvil, Devarsh Patel, Jainish Adesara, Pruthvi Hingu, and Manan Shah. "Exploiting the capabilities of blockchain and machine learning in education." *Augmented Human Research* 6 (2021): 1-14. https://doi.org/10.1007/s41133-020-00039-7
[4] UCI. "A Primer on Machine Learning." *UCI Division of Continuing Education* (2024).
[5] Abroyan, Narek Hovhannes, and Robert Grigor Hakobyan. "A review of the usage of machine learning in real-time systems." *Вестник Национального политехнического университета Армении. Информационные технологии, электроника, радиотехника* 1 (2016): 46-54..
[6] GetEducated. "Pug Dog Earns Online MBA in Rochville University Scam." (2020).

[7] Collier, Lorna. "Top 10 Instant Degree Scams." *GetEducated* (2020).
[8] Gräther, Wolfgang, Sabine Kolvenbach, Rudolf Ruland, Julian Schütte, Christof Torres, and Florian Wendland. "Blockchain for education: lifelong learning passport." In *Proceedings of 1st ERCIM Blockchain workshop 2018*. European Society for Socially Embedded Technologies (EUSSET), 2018.
[9] Li, Rujia, and Yifan Wu. "Blockchain based academic certificate authentication system overview." *IT Innov. Centre, Univ. Birmingham* 8 (2018).
[10] Gopal, Neethu, and Vani V. Prakash. "Survey on blockchain based digital certificate system." *International Research Journal of Engineering and Technology (IRJET)* 5, no. 11 (2018).
[11] Cheng, Jiin-Chiou, Narn-Yih Lee, Chien Chi, and Yi-Hua Chen. "Blockchain and smart contract for digital certificate." In *2018 IEEE international conference on applied system invention (ICASI)*, pp. 1046-1051. IEEE, 2018. https://doi.org/10.1109/ICASI.2018.8394455
[12] Han, Meng, Zhigang Li, Jing He, Dalei Wu, Ying Xie, and Asif Baba. "A novel blockchain-based education records verification solution." In *Proceedings of the 19th annual SIG conference on information technology education*, pp. 178-183. 2018. https://doi.org/10.1145/3241815.3241870
[13] Bhattacharya, Pronaya, Sudeep Tanwar, Umesh Bodkhe, Sudhanshu Tyagi, and Neeraj Kumar. "Bindaas: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications." *IEEE transactions on network science and engineering* 8, no. 2 (2019): 1242-1255. https://doi.org/10.1109/TNSE.2019.2961932
[14] Chelladurai, Usharani, and Seethalakshmi Pandian. "A novel blockchain based electronic health record automation system for healthcare." *Journal of Ambient Intelligence and Humanized Computing* 13, no. 1 (2022): 693-703. https://doi.org/10.1007/s12652-021-03163-3
[15] Shen, Bingqing, Jingzhi Guo, and Yilong Yang. "MedChain: Efficient healthcare data sharing via blockchain." *Applied sciences* 9, no. 6 (2019): 1207. https://doi.org/10.3390/app9061207
[16] Qazdar, Aimad, Brahim Er-Raha, Chihab Cherkaoui, and Driss Mammass. "A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco." *Education and Information Technologies* 24 (2019): 3577-3589. https://doi.org/10.1007/s10639-019-09946-8
[17] Kotsiantis, Sotiris B. "Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades." *Artificial Intelligence Review* 37 (2012): 331-344. https://doi.org/10.1007/s10462-011-9234-x
[18] Siddiqui, Ahsan Ali. "The Epidemiology of COVID-19 Novel Corona Virus in Incidence and the Distribution of the Disease across the World." (2020). https://doi.org/10.34297/AJBSR.2020.09.001407
[19] Lu, Yang. "Blockchain: A survey on functions, applications and open issues." *Journal of Industrial Integration and Management* 3, no. 04 (2018): 1850015. https://doi.org/10.1142/S242486221850015X
[20] Phemex Academy. n.d. "What Are Public, Private, and Permissioned Blockchains."
[21] Seth, Shobhit. "Public, Private, Permissioned Blockchains Compared." *Investopedia*. 2023.
[22] MacDonald, Ross. "Permissionless vs. Permissioned Blockchains: Pros & Cons." *1Kosmos.* 2023.
[23] Binance Academy. "What Are Permissioned and Permissionless Blockchains?" 2023.
[24] Murtaj, Sheikh Mohammad Tahsin, and Abubakar Abubakar Yusif. "PreCogNito GitHub Repository." *GitHub*. 2022.
[25] Breiman, Leo. "Random forests." *Machine learning* 45 (2001): 5-32. https://doi.org/10.1023/A:1010933404324
[26] Benesty, Jacob, Jingdong Chen, Yiteng Huang, and Israel Cohen. "Noise Reduction in Speech Processing." (2009). https://doi.org/10.1007/978-3-642-00296-0_5