Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Integrating Machine Learning and Blockchain for Fraud Prevention in Educational Records

Nagaraju Kasukurthi[1], Dr. Gangadhara Rao Kancharla[2]

[1]Research Scholar, Dept. of Computer Science and Engineering, University College of Sciences, Acharya Nagarjuna University
[2]Professor, Dept. of Computer Science and Engineering, University College of Sciences, Acharya Nagarjuna University
Email: knraju.dwh@gmail.com

Technology is allowing the education industry to advance at a faster rate than in the past, but there are still a lot of uncharted territories and plenty of opportunities for improvement. Machine learning (ML) and blockchain, two of the most game-changing technologies; have contributed to the eroding of more traditional techniques in favour of more technological and effective ones in the field of education. This paper proposes a method that integrates these two light-based radiant technologies to address issues like degree forgeries and other forms of educational record fraud. The basic premise can eliminate the issues of counterfeiting and insecurity surrounding student accomplishments by combining these technologies into a single system that securely stores student data on the blockchain and uses ML to accurately forecast students' future job roles upon graduation.In addition, legitimate data will be trained and predicted using ML models. An official decentralized database of graduation records will be made available to the institution using this method. On top of that, this system gives businesses a place to check their workers' academic credentials. Platforms like LinkedIn, which allows users to manage their professional accounts, provide students with an opportunity to showcase their academic work in an electronic portfolio. Students, businesses, and other sectors may now more readily locate consent for student data.

Keywords: Machine Learning, Blockchain, Traditional Techniques, Graduation Records, Student Data.

1. Introduction

In today's era of rapid technological advancement, many different types of industries are making use of new technologies that provide even more advantages. A good example of how the field of education has shown openness to new technology is the widespread adoption of massive open online courses (MOOCs), which have allowed students to expand their skills and advance in their careers more efficiently. Education may benefit from the use of cutting-edge technologies like machine learning and blockchain (ML), both of which are exhibiting enormous potential and growth in their sectors.Blockchain, the underlying technology of the digital currency Bitcoin, has recently come to the forefront of public attention due to its meteoric surge in value. Its distributed ledger, tamper-proof construction, and openness about accessibility and information are some of its features. Consequently, blockchain technology has been used in several domains, including copyright protection, supply chain management, education, and finance [1]. Blockchain functions similarly to an immutable ledger that records transactions in a certain sequence [2]. Every computer contributes to the total by storing its own linked data set, and new data sets may only be added when the majority of computers in the network agree [3].People are more interested in machine learning and AI because computers can learn from experience and simulate the human brain [4]. A computer program is said to acquire knowledge from experience E in connection with a specific function T and some output measure P if its output on T, as determined by P, grows with E, according to Mitchell, a computer science and ML professor at Carnegie Mellon. With a plethora of training data at their fingertips, several ML algorithms can identify the right pattern, allowing for accurate predictions in these domains [5].

The purpose of this article is to show how machine learning and blockchain may be used in the field of education [6]. To prevent the forgery of student accomplishments and credentials, blockchain technology is being considered [7]. To add to that, it aids in validating student data, which solves the primary issue of educational security. There is a growing body of research on the potential of machine learning (ML) approaches to educational proposal development with the overarching goal of developing tools for mining computational educational settings for useful patterns [8]. A centralised database system is the standard method for storing a student's accomplishments, such as a certificate or degree. The usage of a single credential and the potential breach of a firewall are two issues with this centralised method [9-11]. As a result, without a backup, data might be irretrievably destroyed in the event of database manipulation [12]. Storing student records decentralized, via a blockchain ledger, thereby compounds the aforementioned firewall issue [13]. Due to the availability of many copies of the blockchain, data cannot be lost.Furthermore, some educational institutions keep student records on an inaccessible local system, making it more difficult to remotely access the data and increasing the likelihood of forgery due to the lack of official documents available for remote cross-verification. Using blockchain technology, remote access to student data becomes a breeze [14].

This article proposes a system, and here's how it works. The first step is for the school or institution to submit the student's profile, which includes information about their grades, extracurriculars, degree, and any other certifications of accomplishment they may have from the course.Afterwards, a blockchain will be constructed using these details [15]. The ML algorithms fed this blockchain as a trustworthy data source, and the students will be guided into a career path that better suits their interests and abilities. Another perk is that throughout the hiring process, companies may trust the data kept on the blockchain, which means that applicants have no opportunity to present fake papers.

Two crucial aspects of the proposed methods are covered in section 2: (i) the operation of the system and (ii) validity operations. Section 3 of the paper contains the results and an analysis of the suggested system. Difficulties and potential uses are given in Section 4 which also delves into the difficulties encountered during system development and the potential future uses of the system. Section 5 of the Conclusions section provides some last thoughts on the suggested system's use of two blossoming technologies in tandem.

2. Methodology

The immutable, distributed, and transparent ledger known as blockchain is attracting interest from educational institutions as a safe way to keep students' academic data. This central, immutable database makes it easy to quickly and easily check a student's credentials and stops document forgeries.The immutable, distributed, and transparent ledger known as blockchain is attracting interest from educational institutions as a potentially safe way to keep students' academic data. This central, immutable database makes it easy to quickly and easily check a student's credentials and stops document forgeries.

The suggested solution aims to improve education by combining blockchain and ML capabilities. Students, employers, colleges and universities, and authorities are all parties involved in the management and access to educational data, and the authors have taken their needs into account to design a secure system, simple to use, and always available. Figure 1 depicts the proposed system stakeholders and their duties.

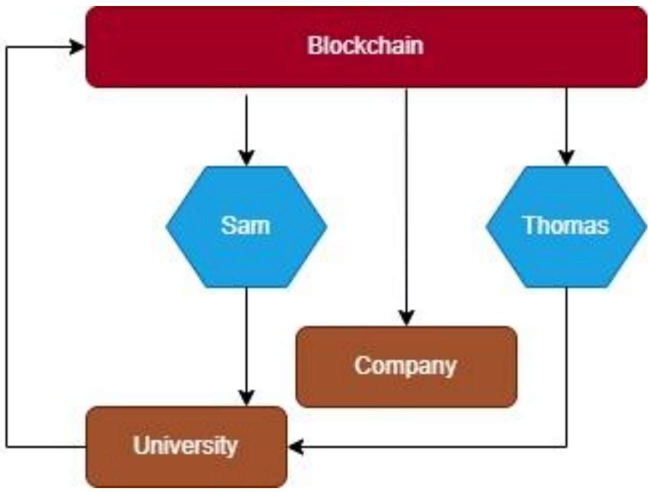

Figure 1 Proposed System Stakeholders and their Duties

The proliferation of forgeries and fraudulent degrees necessitates the use of blockchain technology at the academic level. To keep track of all of the alumni, the institution needs to implement a standard, decentralized, and authenticated database. To ensure the system's trustworthiness, no outside force should be able to manipulate or erase this academic data. Once students have passed their final exams, their information should be added to this distributed database. This information should include their grades in their major classes (e.g., computer engineering), their final cumulative grade point average (CGPA), the amount of work they have done throughout the course, any extracurricular activities (e.g., online quizzes), and any athletic accomplishments they have had.There should be no tampering with the data, and it must be maintained securely to prevent loss. Since blockchain is a distributed ledger, there may be many copies of the data; hence, if one data chain is compromised, other data chains may be accessed. An example of academic data that fits this description is a block of student registration numbers uploaded to a blockchain, which may help to trace the ownership of any digital asset transacted between two anonymous parties. Accountability, anonymity, and the avoidance of fraud are three essential aspects of blockchain-supported security.

The following are the most basic authentication capabilities of a blockchain.

(1) Record: Every transaction is recorded in the blockchain's ledger. The data in the ledger cannot be changed since it is a blockchain. Every node has received the ledger.

(2) Blockchain: The hash value of the preceding block is included in each block, creating a chain. Data corrections in one block (n) modify the hash value but do not check the hash in the block after it (n+1). This is a chain reaction that affects the chain as a whole. This quality enhances the protection of sensitive information and expertise.

(3) Privacy: By limiting database users to seeing only authorized transactions, blockchain technology guarantees privacy.

(4) Transparency: Blockchain allows for the transparent storage and handling of transactions and the ledger state by sharing the ledger with all nodes and utilizing consensus procedures to establish consensus among all nodes. Frequently, consensus methods ensure the sequencing and completion of transactions.

(5) Cryptography: By using hash-based methods that provide a set hash depending on the block's content, cryptography enables encrypted transactions and makes the blockchain immutable.

(6) Smart contracts: These are essentially contracts that are built into blockchain technology and include a set of rules that govern how the parties involved may interact with one another. The plan is carried out mechanically if the specified criteria are satisfied. Without the agreement of the whole network, no deal may be carried out.

The system's security has been guaranteed up to this point. The system's second component was machine learning (ML). As covered in the previous part, machine learning has several advantages in the educational system. If the data is fed into a method for prediction is invalid and fraudulent. Moreover, if the data used to train the model has not undergone validation. In such instances, just a trash-in garbage-out procedure will be used. The primary objective of the proposed system is to ensure the provision of accurate and verified predictions by supplying the ML model with reliable data.The data is kept on the blockchain in such a way that it cannot be accessed or altered by any other party. Furthermore, the system was designed to prevent the loss of students' educational records. Predictions on the differentkinds of jobs that students should get upon graduation will now be made. Using this method, a student may get reliable career advice based on their academic history. In this case, the ML model is trained and trained using student data directly extracted from the blockchain.

Employers stand to gain the most from this system, which is its third component. The veracity of the employee's records will no longer be an issue for these businesses. To validate the identities of graduating students, trusted companies may access the university's blockchain. As a result, these businesses will have to include blockchain data verification in their hiring process. The next generation of students also has the option to use any reliable source, such as the professional identity management network LinkedIn, to disseminate their information. This information will be considered the pupils' official record. As seen before, Figure 2 depicts the workflow of the suggested system.

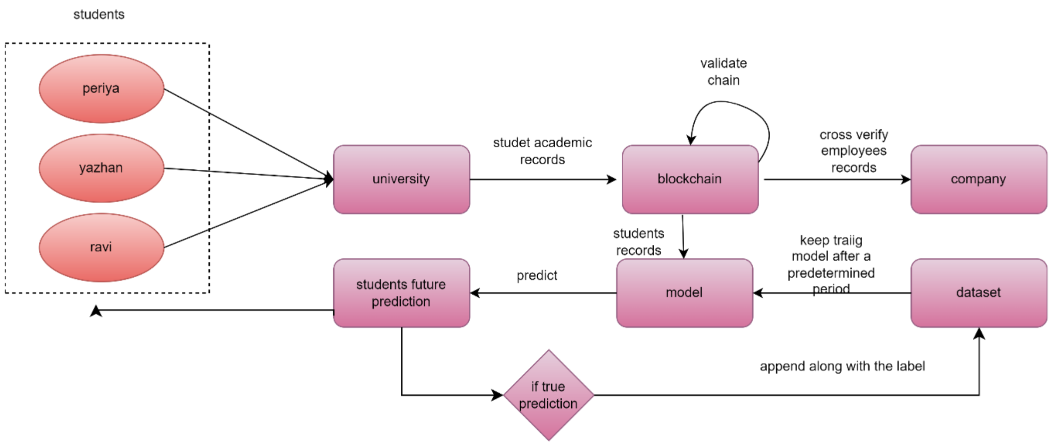

Figure 2 Workflow of the Suggested System

This system is described hierarchically so that the stakeholders and their duties can be better understood. Each level has a distinct collection of stakeholders and their respective tasks. Figure 3 illustrates the suggested system's hierarchical structure.

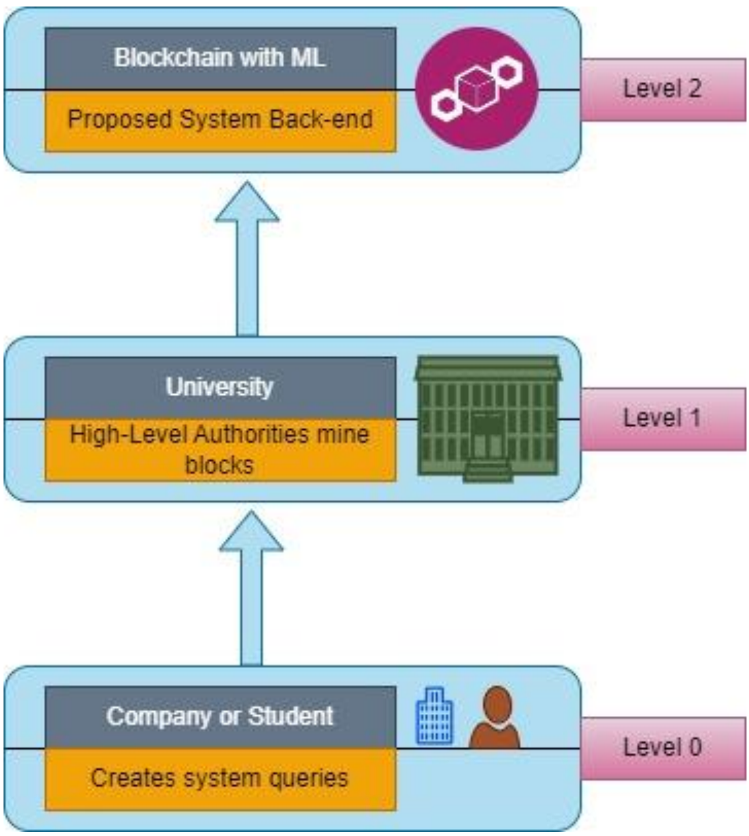

Figure 3 Suggested System's Hierarchical Structure

2.1 User (0-level)

The user level is named because it provides details about the system's interfacing object. To make the system more user-friendly, it is helpful to learn the user's demographics and preferences via this categorization of interacting entities. Because it categorises the system's users as students, businesses, or other organizations, this helps in defining the correct structure of the system. Here, businesses and students alike may run queries against the system to access the blockchain's recorded data.

2.2 Level 1: Intermediate

The authority that oversees the system and acts as a mediator within it is described at this level. An intermediate-level team, that is, the administrative authority, is responsible for mining the student data block and calculating its proof of work. This group is also responsible for keeping the system running smoothly and fixing any problems that may arise. Consequently, the intermediate level serves as the glue that ensures the user remains glued to the system.

2.3 Level 2: Processing

A decentralized database is located on this level. The processing of the user-fired query takes place at this level. This layer processes the user's request and returns the result. If this layer fails, the whole system can come to a standstill since it is the most important component. Consequently, it requires the greatest caution when handled. At this level, there are several kinds of backend software that service, including the Ethereum virtual machine.

3. Results and Discussion

Three crucial aspects are included in this section to assist users in navigating the proposed system. The first section of the suggested system is the machine learning section, which details the dataset and the models' respective accuracy levels. The second part explains the system's blockchain component, which walks through the proposed system's blockchain structure.Lastly, the third and most crucial paragraph explains how the two technologies are integrated.

3.1 Machine Learning Stage

Gathering domain-specific data is the first stage of every ML-based research. To compile this dataset, a Google form was used to ask a series of questions on the academic backgrounds of 1,550 students who received degrees from the Computer Department at Government Engineering College in India. The dataset includes information like students' marks in computer engineering courses, which are rated from 1 to 100.After the survey was over, the data set included one dependent feature with categorical values and twenty independent characteristics. It includes the student's cumulative grade point averages (CGPAs) together with other information including their marks from the 10th and 12th grades, their involvement in technical quizzes and projects, their rank, their gender, the number of backlogs they have, various technical events they have participated in, and their number of athletic successes. The dataset also includes attributes that indicate the name of the board the student attended for 10th and 12th standard. The encoding of labels or one hot encoding, in conjunction with a specific ML method, is required to further reduce the five category characteristics included in the dataset to numerical features.

Furthermore, out of the total number of numerical characteristics, eight are discrete and the rest are continuous.In this dataset, the student's work serves as the dependent feature. This feature is reliant on the other characteristics; it is the goal feature. Data that may be sourced from a student's official university records is taken into account while creating the dataset.

The exploratory data analysis in Figure 4 depicts the feature relevance as an X-axis "feature score" that simply sums up the number of times each function is divided. Among the dataset's most salient features is the normalized rank. The CGPA's position as the dataset's second most significant attribute sheds light on its impact on the job role. See a seaborn barplot (role in the company vs. CGPA) in Figure 5. To illustrate the range of possible values for each bar when mistakes are included, an error bar is included in the bar chart. The error bars in the following graphic are based on the median aggregate.

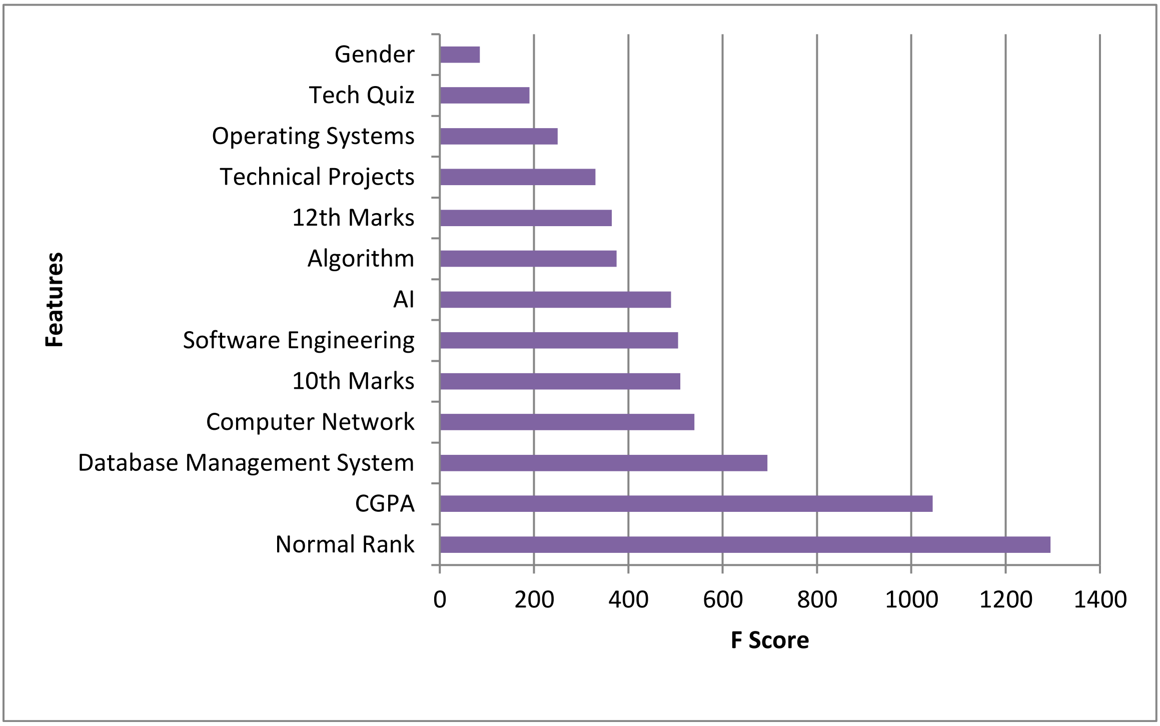

Figure 4 Exploratory Data Analysis (Feature Relevance)

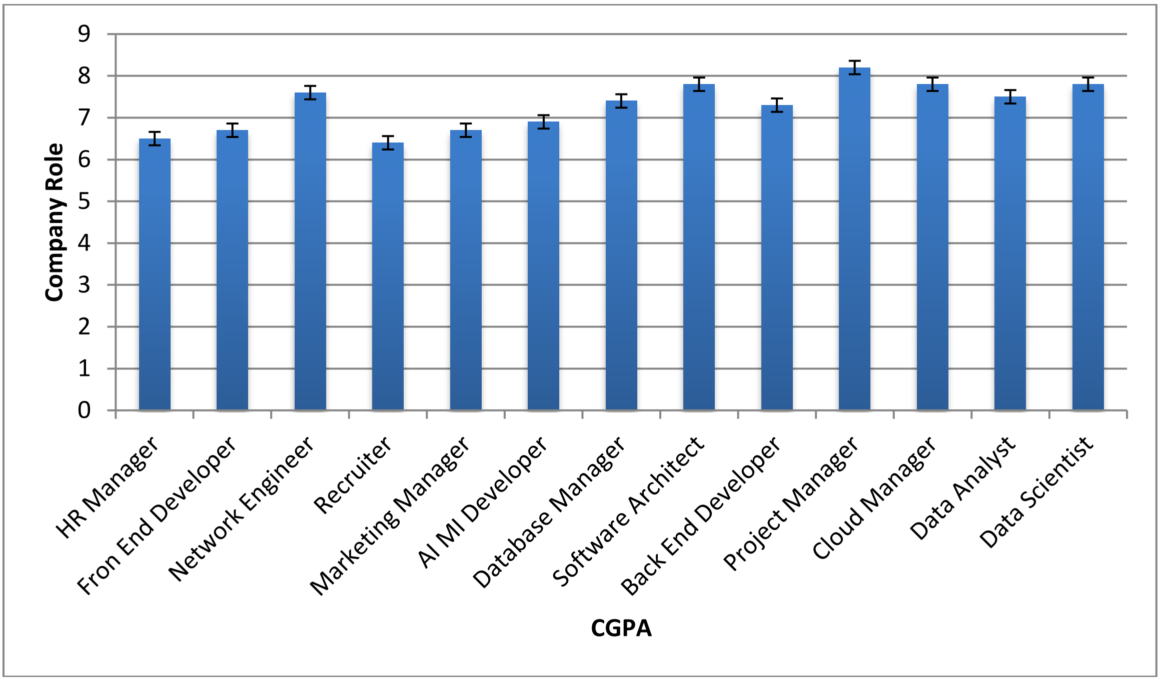

Figure 5 Seaborn Barplot (Role in the Company vs. CGPA)

Figure 6 shows the seaborn violin plot, which helps to understand the distribution of students' grades in the dataset. Since violin charts provide an estimate of the dataset rather than mapping each data point individually, they are ideal for depicting data distributions. A pair of parallel kernel density estimator charts runs the length of the violin plots. The thick line in the centre of each violin denotes the interquartile range, while the white dot on the line marks the median. A 96% confidence interval was shown by the line that extended from the thickest portion to the ends.

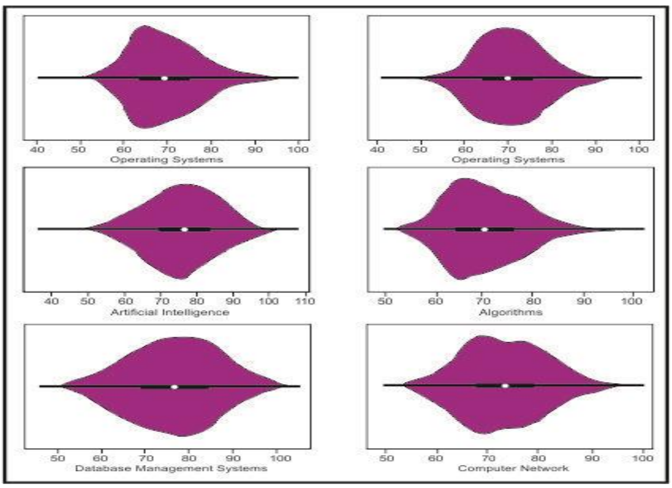

Figure 6 Seaborn Violin Plot

Furthermore, Figure 7 displays the accuracy of the various models evaluated for the datasets used to predict the employment position based on the student's academic records. A large number of ML classifiers were also tested on the dataset.

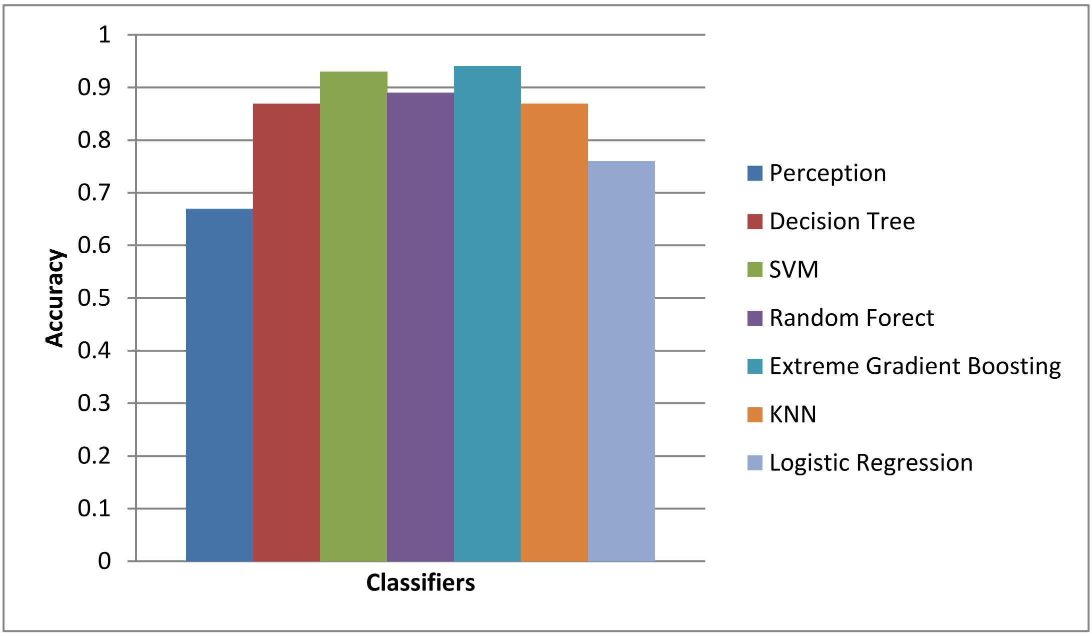

Figure 7 Accuracy of the Various Models

The two models that performed the best out of all of them were the support vector machine and XGBoost. A synopsis of the two models is provided below: (i) Machines that provide support: This robust ML technique determines the class of an unclassified point by first creating decision boundaries (sometimes referred to as a "decision hyperplane" in this context) across classes. (ii) XGBoost: An ensemble learning boosting approach. The results it produces are unparalleled by any other model that has been trained using the boosting approach.

3.2 Blockchain Stage

Python classes are used to represent blockchains and blocks, including initialized methods and variables. Blockchain class methods include inserting blocks, verifying chains, and calculating proof of work. Python's safe hash algorithm 256 is used to compute the block class's hash.Algorithm 1 displays a code sample that demonstrates a straightforward approach for appending a block to a chain. This approach involves passing the operations, which in this scenario represent the student data, as a Python dictionary. Subsequently, a new block is generated and added to the chain.

Algorithm 1 Code Sample

```
def add-block (self, transactions):
#get previous hash
previous_hash=(self.chain[len(self.chain)-1]).hash

#create a block
New_block=Block(transactions,previous_hash, str(datetime.datetime.now()))

#compute proof-of-work
proof=self.proof_of_work(new_block)

#add it to the chain
self.chain.apend (new_block)

return proof, new_block
```

3.3 Integrating Blockchain and ML

The API contains a code fragment that responds to prediction queries. The blockchain is executed on the Flask server, which is where the request is sent. Consequently, the ML model receives its data from the server-side blockchain. A reference to the blockchain class is made using the variable "blockchain" in the code below. The steps outlined before were followed to add the blocks. A dictionary-style 'transactions' variable stores the student data.The information about the pupils was taken from the database and put into the "student_data" table. To get predictions, this list solely comprises feature values. Predictions were made using the top two models, XGBoost and a support vector machine, and the categorical features were managed with the help of a label encoder.

4. Difficulties and Potential Uses in the Future

First, there is room for expansion in the study of blockchain technology as it pertains to education. For blockchain technology to make a significant difference in the field of education, a fundamental shift is required, along with cross-sector collaboration from all relevant parties. As a methodologically sound strategy, research on specific student behaviours and their subsequent achievement is essential for educators. It is quite challenging to evaluate this kind of educational venture using a pre-set smart system without human intervention.A few schools are wary about giving blockchain access to their data, or they don't know what data to provide because of the uncertainty around the blockchain. It is a huge undertaking to compile blockchain records from different institutions.

Second, getting a group of students by their enrolment number may be a major problem with this technique. Python dictionaries were used by the authors to create the chain, which is then added to a Python list. For the suggested system to use a student's course grades to forecast their future employment, it is necessary to get the student's block from the chain. Using a for-loop, the suggested system iteratively checks each block's enrolment number against the one from which the data is now desired. Theextracted data is sent into the Support Vector Machine (SVM) and XGBoost algorithms if a connection is found. This assignment has a temporal complexity of O (n), wherein n is the total length of the chain structure. Even while only using a small number of blocks to illustrate the system, in practice, the length of the chain might grow exponentially as each year brings more graduating students.Parallel programming is one alternative that may be used to effectively loop over the chain. Also, as mentioned in the methodology, if the ML model's predictions come true, the data and label may be added to the training dataset. The whole picture of the system under discussion may be adequately portrayed by doing a SWOT (strengths, weaknesses, opportunities, and threats) analysis.

Figure 8 depicts the suggested system's SWOT analysis, which provides a synopsis of the whole system including the several elements that allow for a better recognition and conclusion about the system. The purpose of the SWOT analysis is to outline the system's benefits, drawbacks, and overall scope.

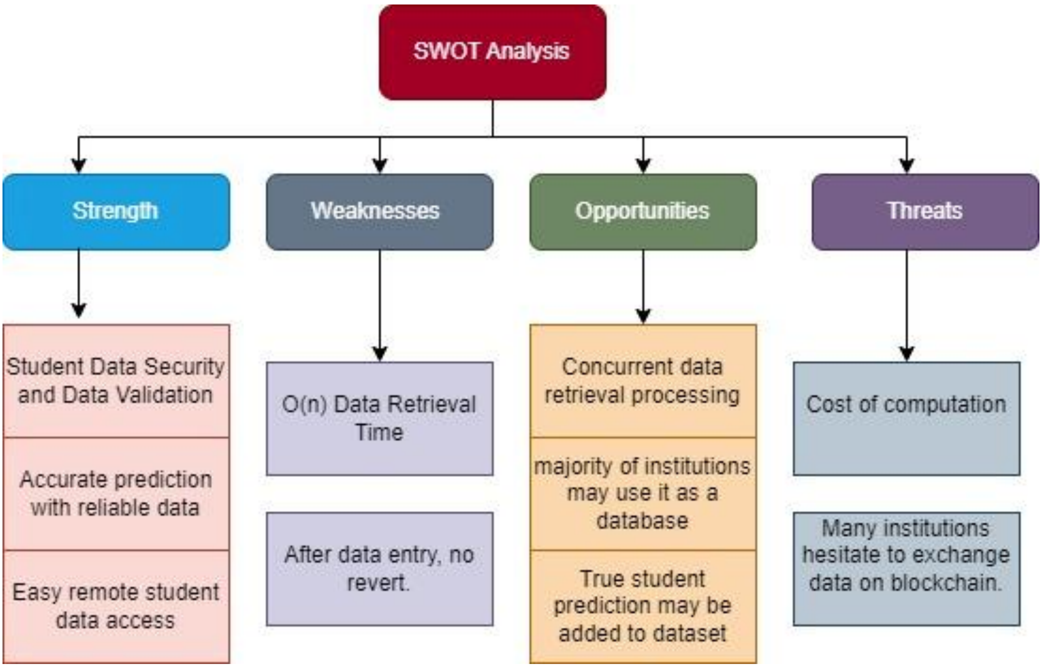

Figure 8 SWOT Analysis

5. Conclusion

Examining the different technologies employed in education to improve student outcomes, it was found that the application of ML gives students predictions about the happenings in the future and that blockchain technology has proven effective in reducing the possibility of student accomplishments being forged. Thus, a system that integrates blockchain with machine learning was covered in this study, giving the user the advantages of both methodologies.Another way to make ML-based models more reliable is to use them on a trustworthy dataset that has been verified using blockchain. This study the process by which a university verifies student work by mining their blocks. To improve its prediction accuracy, this ML model may use trustworthy data as input. Therefore, this paper's main objective is to depict a system with a blockchain that stores student data in a decentralized database (ledger). The idea is that a job placement service can query this database directly, get the student's verified and validated data very fast, and then use a trained ML model to make accurate predictions about the student's suggested job.

References

1. Kumar, K. Sunil Ratna, et al. "Iot and data mining techniques to detect and regulate of the solar power system." 2023 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2023.
2. Bello, Halima Oluwabunmi, Courage Idemudia, and Toluwalase Vanessa Iyelolu. "Integrating machine learning and blockchain: Conceptual frameworks for real-time fraud detection and prevention." World Journal of Advanced Research and Reviews 23.1 (2024): 056-068.
3. Shah, Dhruvil, et al. "Integrating machine learning and blockchain to develop a system to veto the forgeries and provide efficient results in the education sector." Visual computing for industry, biomedicine, and art 4 (2021): 1-13.
4. Ashfaq, Tehreem, et al. "A machine learning and blockchain-based efficient fraud detection mechanism." Sensors 22.19 (2022): 7162.
5. Hisham, Sabri, Mokhairi Makhtar, and Azwa Abdul Aziz. "Combining multiple classifiers using ensemble method for anomaly detection in blockchain networks: A comprehensive review." International Journal of Advanced Computer Science and Applications 13.8 (2022).
6. Pranto, Tahmid Hasan, et al. "Blockchain and machine learning for fraud detection: A privacy-preserving and adaptive incentive-based approach." IEEE Access 10 (2022): 87115-87134.
7. Mohammed, Mohammed A., Manel Boujelben, and Mohamed Abid. "A novel approach for fraud detection in blockchain-based healthcare networks using machine learning." Future Internet 15.8 (2023): 250.
8. Tyagi, Amit Kumar, S. U. Aswathy, and Ajith Abraham. "Integrating blockchain technology and artificial intelligence: Synergies perspectives challenges and research directions." Journal of Information Assurance and Security 15.5 (2020): 1554.
9. Al Wahaibi, Aisha Abdullah Ali, and Manju Jose. "Merging Artificial Intelligence & Blockchain Technologies to Solve Academic Qualification Forgery Issues." Journal of Student Research (2019).
10. Bhandari, Aradhita, Aswani Kumar Cherukuri, and Firuz Kamalov. "Machine learning and blockchain integration for security applications." Big Data Analytics and Intelligent Systems for Cyber Threat Intelligence. River Publishers, 2023. 129-173.
11. Amponsah, Anokye Acheampong, Adebayo Felix Adekoya, and Benjamin AsubamWeyori. "A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology." Decision Analytics Journal 4 (2022): 100122.
12. Tang, Qiang. "Towards using blockchain technology to prevent diploma fraud." IEEE Access 9 (2021): 168678-168688.
13. Arza, Mahita Sri, and Sandeep Kumar Panda. "An integration of blockchain and machine learning into the health care system." Machine Learning Adoption in Blockchain-Based Intelligent Manufacturing. CRC Press, 2022. 33-58.
14. Senthilselvi, A., Aniket Kumar, and Manoharan Charan Kumar. "A Novel Approach to Carrier Guidance System using Machine Learning and Blockchain." 2024 International Conference on Computing and Data Science (ICCDS). IEEE, 2024.
15. Zhang, Guoming, et al. "Identifying fraud in medical insurance based on blockchain and deep

learning." Future Generation Computer Systems 130 (2022): 140-154.

16. Boughaci, Dalila, and Abdullah AK Alkhawaldeh. "Enhancing the security of financial transactions in Blockchain by using machine learning techniques: towards a sophisticated security tool for banking and finance." 2020 first international conference of smart systems and emerging technologies (SMARTTECH). IEEE, 2020.
17. Shaikh, Zaffar Ahmed, et al. "Blockchain hyperledger with non-linear machine learning: A novel and secure educational accreditation registration and distributed ledger preservation architecture." Applied Sciences 12.5 (2022): 2534.
18. Kamišalić, Aida, Renata Kramberger, and Iztok Fister Jr. "Synergy of blockchain technology and data mining techniques for anomaly detection." Applied Sciences 11.17 (2021): 7987.
19. Shrivastava, Vineet, and Suresh Kumar. "Utilizing blockchain technology in various application areas of machine learning." 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon). IEEE, 2019.
20. Joshi, Kunj, et al. "Machine-learning techniques for predicting phishing attacks in blockchain networks: A comparative study." Algorithms 16.8 (2023): 366.