# A Blockchain-Based Document Verification System for Employers

## GitHub Repo: https://github.com/wimpywarlord/ISM-PROJECT

## 1.Abstract

Academic Institutions often maintain records and documents of the students enrolled within the institute. These records are greatly regarded by employers and often verified before giving employment offers. While all the other sectors in academia are moving towards automation, this specific use case is still quite primitive in its functioning. Even when this is a regular activity undergone by academic institutes, methodologies with vulnerabilities and reasonable overhead are employed. There is a need for a tamper-proof, reliable, and faster mode of document verification for employers which can be used on the go with complete reassurance for the authenticity of the provided documents. This paper puts forward a blockchain-based platform that allows academic institutes to offer a secure, dependable, cost-effective, and scalable verification of documents via a web interface. The platform is built upon the logic written in solidity and utilizes the Ethereum blockchain to enforce immutability of the document, the backend is linked to the user interface using the web3.js library. Additionally, the platform offers a statistical comparison of these records to other students and assists the employer to assess the performance of the student being recruited.

## 2. Introduction

Most processes in the modern world have started using automation and digital architectures. This has improved efficiency and reduced the cost and time for these processes. Employers require academic institutions to share documents for verification purposes when a student is joining their organization after being recruited on the campus of the institution. The employers always verified due to the possibility of false documents being presented by the students. Hence as an authenticity check, they ask the academic institution for the record of the hired student. In most cases, the transfer of academic documents and records is done using primitive methods like postal mail and e-mails and then they need to be verified manually. Postal mails require a person to physically send and receive the mail. This is followed by a manual verification procedure. This makes postal mails a time taking, costly, and less reliable method for such important tasks. E-mails require a person to physically scan the documents, send the e-mail and it depends on the person on the other side to check the documents and verify them. It is prone to problems like entering the wrong e-mail address, e-mail going into the spam folder, or sender attaching wrong attachments.

Considering the drawbacks of the above-mentioned methods, machine learning algorithms like SVM have been used for identifying whether the documents are original or not [1]. Also, various institutions have designed web interfaces where the documents are uploaded by the concerned authorities, verified by the system using digital signatures, or QR code scanners, or using hashing algorithms. These systems comprise of external universities, recruiters, students, and the current universities who play their respective roles for the smooth functioning of the system.

These approaches provide some benefits over the primitive method of postal mail and e-mails but still have their limitations. For few models, the type of documents to be uploaded was limited to PDF format only and other text documents and image-based documents were not acceptable [3]. The main issue with QR code scanners was that it is not cost-effective and faking QR codes and digital signatures have a high possibility [4]. Hacking and data getting deleted from DB are also possible when it comes to web interfaces.

To resolve many of these limitations a blockchain-based approach can be used. Blockchain is a decentralized, immutable digital ledger that can be used to store information. Being an immutable ledger, once a verified document is put into the blockchain, the trustworthiness of the document being genuine is guaranteed. A blockchain-based approach provides full immunity from situations such as a student's record being deleted by error or someone hacking into a database to change records. The

information of the students will be available to employers on-demand over the blockchain and they can be sure about the data being un-tampered and verified by the academic institution. Most of the steps in this process can be automated and won't require the interference of a human. The purpose of our proposed system is to provide a user-friendly, fast, reliable, and secure system for the verification of academic records.

The structure of the rest of the paper is: Section II is about related work, Section III is the background which explains the working of blockchain, in Section IV we present our proposed solution, Section V is about evaluation and results of the experiments and the finally Section VI presents the conclusion and expected future work.

## 3. Literature Review

There have been various possible solutions for the transfer and the validation of academic records and self-sovereign documents which have been adopted by various institutions and individuals worldwide for various purposes varying from academia to healthcare to legal deeds.

Utpal Garain and Biswajit Halder [1] proposed a model which uses both genuine and duplicate bank cheques as the reference which are scanned into colour images and security features are then extracted to generate feature vectors. A classification model is designed using SVMs which is trained using a set of labelled samples and a testing set is used to determine accuracy. The analysis of error cases is not discussed in the paper. Also, this approach is limited to bank cheques only and needs to be extended to lottery tickets, legal documents, etc.

HS Brdesee [2] from King Abdulaziz University proposed a unified e-portal for the university to standardize the verification of all the documents of the students when the recruiters and other universities demand it. The recruiters, external universities, students need to log in and then create a verification request or trace the previous request. They also get to choose the type of document to be verified and the means by which they wish to receive the verification.

IDStack provides a platform [3] where the documents provided by the owners are digitized by the extractor and converted into a machine-readable document using IDStack stating that the provided documents and the digitized document are 100% matching. Each machine-readable document is then digitally signed by the validators. Those signed documents are then viewed by relying upon parties and confidence and correlation scores are provided to the document. One major issue with digital signatures is that these documents are self-signed and there is no way to verify the identity of the signer. Another drawback is that the model only accepts PDF documents and not other text documents and image-based documents.

A possible update to this technique would be supporting image processing and using OCR to extract data from image files.

Putro, P. A. W.and Luthfi, M [4] provide a system to produce authentic and safe documents with a combination of perceptual hash and OCR. Public and private keys are determined to be used in both, document creation and verification phases. SHA-256 is used to hash the contents in the document and a digital signature is created using hash value and keys. This along with perceptual hash and OCR hash is used to form QR code which will be inserted in the document to be printed. The QR code reader will read the QR code given in the document and test its authenticity and integrity. The major issue with this scheme is that devices used to scan QR codes are costly and faking a QR code is also a possibility nowadays.

The above methods are very well tested and are quite efficient for mainly document authentication and transfer if required while blockchain-based approaches provide extra competence to achieve immutability through its distributed continuous ledger architecture.

The MIT Digital Credentials Consortium [5] provides a model where the learner can register their digital documents to a bitcoin blockchain. The learner is issued digital credentials which they can associate with various self-sovereign IDs. The credentials are under the learner's control and the decision of sharing or exchanging them depends on him/her. These credentials once shared with the parties, are verified and the status of verification is returned in the form of success or failure with the reason behind the failure. Few drawbacks of bitcoin blockchain include very high processing time, immutability, harder to scale and high energy consumption.

The ShoCard app [6] which is designed to prove one,s identity rather than verify someone's identity, uses blockchain technology as a public, unchangeable ledger that allows third parties to verify that the original data or certifications have not been tampered with. It provides a wallet to store all validated identities such as driver's license, school transcripts, etc. Every user will be provided with a credit score, and based on certain conditions, the identities will be validated. Temporary access to the private sections of blockchain can be provided to few organizations for validation.

EduCTX [7] which is a blockchain-based higher education credit and grading platform, provides an approach that aims to form single merit for judgment for all students/employees. It is based on a P2P network system. Home universities referred to as HEI is also a part of the blockchain network. When the student enrolls in HEI, a student ID is issued and a new blockchain address is generated. At the point when an organization needs to confirm the students' credit record and course completion, the student needs to send his/her blockchain address, 2-2 multi-signature blockchain deliver and reclaim content to the verifier - organization. Utilizing the blockchain web API to get to blockchain information, the organization checks the measure of ECTX tokens in the 2-2 multi-signature address, which addresses the students' credit count.

Permissioned blockchain-based system [8] is an approach for the verification of academic records. It consists of a web interface for registering and demanding the transfer of academic records with a backend utilizing hyper ledger fabric and hyper ledger composer to preserve the hash of the archives on the blockchain for validation. The usage of blockchains such as ethereum and bitcoin function on open networks and are therefore not trustworthy and are unsafe. This calls for the introduction of permissioned blockchain such as hyper ledger with functions on a private network and the users can be identified and hence makes the functioning safe and trustworthy. As compared to the other approaches discussed above, this model provides flexibility to upload documents in any format and it provides an automated solution that is more scalable to assist the load of possible users and institutions.

## 4. Proposed Model

This section attempts to explain the proposed solution of academic and employment document verification systems powered by blockchain. To best grasp the overview of the proposed technology, we will start by understanding the underlying assumptions which were axiomatically assumed before building the platform, followed by the description of the architecture of the system. Then the implementation process is discussed along with the scope of scalability. Finally, some challenges and solutions are discussed which were encountered during the development lifecycle.

A. Assumptions and Scope

Some non-trivial assumptions are to be made, for the proper functioning of the platform.

1. It is assumed that the documents, transcripts, student information and credentials, history of academic performance, and other such details are provided exhaustively by the participating university/entity/organization.

2. It is assumed that the documents and details provided by the participating organization are accurate and true. The integrity of the details provided by the participating organization during the system initialization is attested by the organization itself.

3. The records and details provided to the system are valid for the indefinite foreseeable future unless particularly specified by the participating organization, in which case it will be explicitly mentioned on the interface.

The proposed technology offers a comprehensive solution to documentation and detail verification for academic institutes and employment houses, provided that the above assumptions are met.

B. Architecture

The system architecture can be broken down into three main modules. Namely the (1), smart contract, the actual (2) ethereum private block-chain network, and finally the (3) web interface module. In the rest of this subsection, we will start to discuss each module, for the best comprehension.

Ethereum Smart Contracts are a methodology developed and provided by the ethereum blockchain community, which allows any blockchain idea to become highly scalable. This is achieved by introducing a programming language called solidity [19], where the logic to be followed by the blockchain is specified. With this algorithmic set of rules written in solidity the associated ethereum blockchain network knows exactly how to function. The ethereum community additionally offers an integrated development environment (IDE) called the Remix for the solidity programming language, which offers features such as solidity linting, syntactical checking, and easy compilation of the code. For defining the data within the blockchain, *structs* data types are most popularly used. For our proposed technology, the logic for the document verification system is written in solidity, where the struct data type defines the attributes to be stored for the student, all within the Remix IDE.

Ethereum blockchain network is the second layer of the proposed architecture and it is essentially a network of nodes/participating devices working together to validate transactions and maintain a ledger based on the norms of the ethereum community. There are multiple ethereum blockchain networks within the community, for different purposes. This blockchain network is where the data of the students will be stored once the smart contract is deployed. Once the contract has been deployed and the data has been pushed and validated by the network, it fundamentally becomes tamper-proof. The deployment is done using the Remix IDE.

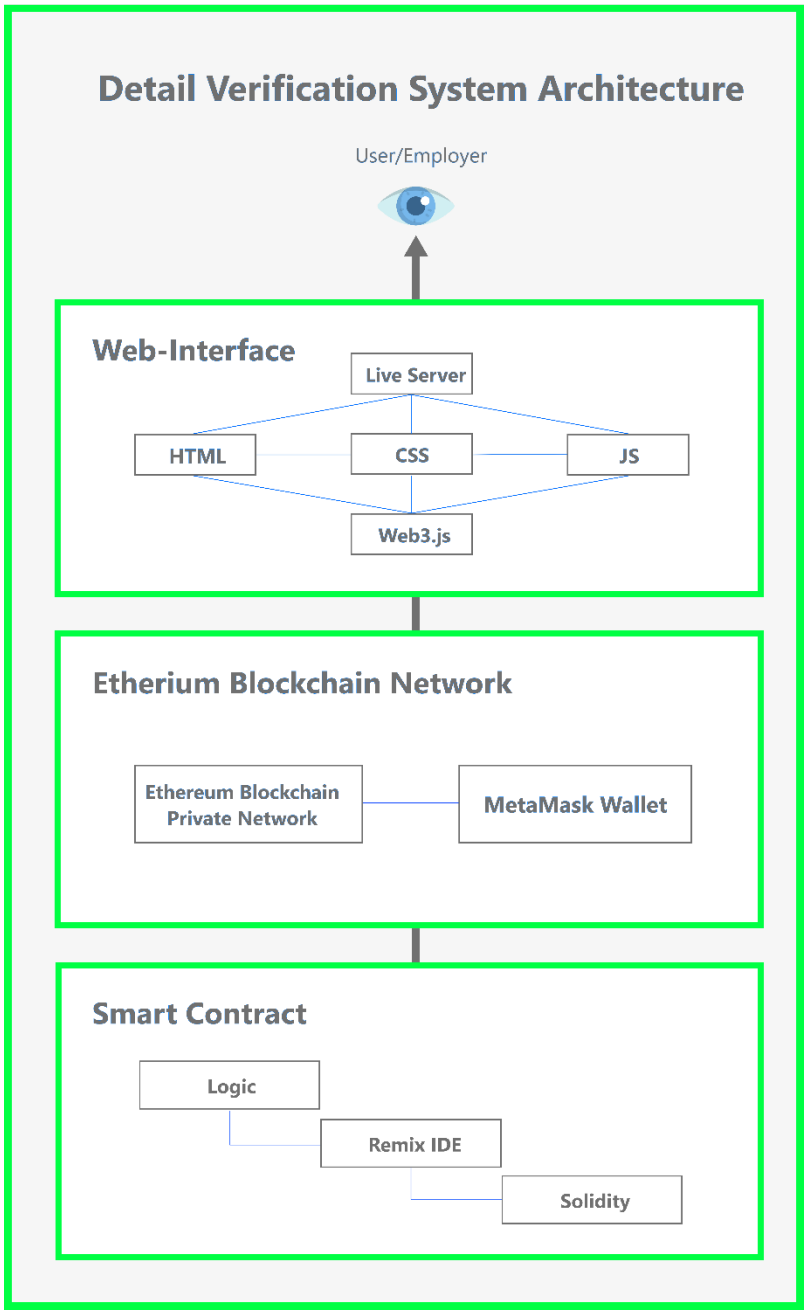


**Figure 1: System architecture for the proposed technology.**

The last and the topmost layer of the proposed architecture is the web interface, where the employer or academic authorizes can come to verify information and documents. This web interface is connected to the blockchain using a Javascript framework- Web3.js. This web interface abstracts and encapsulates the other two layers, and makes them invisible to the user. The user simply searches the desired candidate's information and verifies it.

C. Implementation

As discussed in the previous section, the proposed technology was implemented through three modules. This section will elaborate more on each of these 3 modules and their implementation in real-time.

Smart Contract Module

The smart contract is written in the solidity programming language, which is considered to be a mild representation of C++ and Javascript. The smart contract is essentially the logic behind the functioning of the blockchain. For our particular use case, our smart contracts implement the logic behind document and detail verification. The solidity program a..k.a the smart contract defines functions within the code to perform functions such as creating a student, and reading student details with functions *createStudent()* and *readStudent()* respectively. The details within each student's block are determined by defining a *struct*, which is essentially a user-defined data type that helps represent complex real-world entities in the code easily. This contract is written within the Remix integrated development environment which facilitates solidity development. The smart contract for our proposed application is written in the solidity version *0.5.0* and is further augmented with *abicoder v2* which allows us to use structs and dynamic variables along with functions. The code for the smart contract can be found and downloaded at https://github.com/wimpywarlord/ISM-PROJECT.git

Ethereum Blockchain Network

The ethereum blockchain network used for the proposed technology is *Rinkby* private network [20]. This network is a typical ethereum network that utilizes the Proof of Work (PoW) algorithm to verify transactions to and hence limits the user from tampering with the data. In order to deploy the smart contract to the rinkby network, the *metamask* cryptocurrency wallet is used to pay the gas money [21] as shown in figure 2.

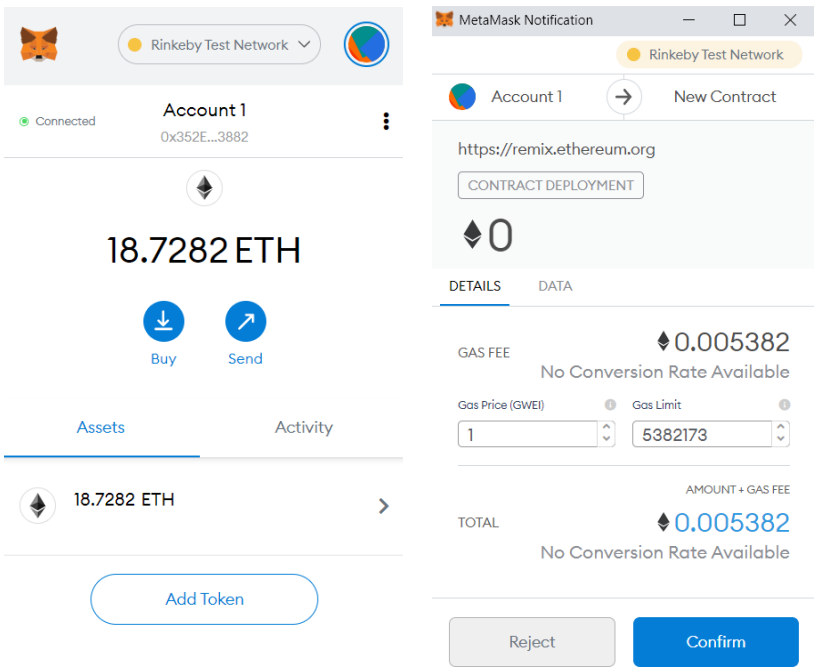


**Figure 2: Metamask cryptocurrency wallet being used for payment of gas fee while the deployment of smart contract on the ethereum Rinkby test network.**

To attain ether in the rinkby network wallet, public faucets were used. Upon linking the metamask wallet with the Remix IDE, the contract is deployed on the network. With this, the blockchain is set up and additionally equipped with the smart contract which is the logic of functioning.

Web Interface

The web interface is the topmost layer of the whole architecture. This is the only layer that is visible to the user, other layers are of no significance for the user. The web interface needs to be connected

with the blockchain. This is done using the *Javascript* framework *Web3.js* [22], which is developed to enable interactions between ethereum networks and nodes over *HTTP/IPC/WebSocket* protocols. Additionally, the *Application Binary Interface* (ABI) for the smart contract [23], along with the deployed contract *to address* is fed to the web3.js function, to perfectly establish the connection. Once connect the interface is structured using HTML, styled using CSS, and manipulated using Javascript.

Scalability

The platform is built upon the ethereum blockchain and hence scaling is a very straightforward process. As per the current standard the chain can store up to 13TB of data. Additionally, if the data requirements increase, a simple enhancement of memory of the full nodes will vastly increase the ledger capacity. Moreover, the simple web interface for the users, allows them to use the service without being connected to the ethereum blockchain directly. Therefore, even in the state of storage issues, the platform is fully functional.

Challenges and Solutions

Currently, the ethereum blockchain can support only up to 30 transactions per second which acts as a bottleneck in some edge cases where rapid update queries in data are issued for multiple participating organizations at the same time. However, the *ethereum 2.0* is expected to be released by the end of this year, which will push the threshold to around 100000 transactions per second.

## 5. Plan for review 3

We implemented the following features for our review 3 presentation.

1) Login for extra layer of Security.
2) Skill Filtering Feature.
3) Transcript.
4) More types of Graphs for better understanding and contrast.
5) Hosting and attaching of domain.
6) Creators Page.
7) Not found Page.
8) Form validation in all input fields.

## 6. Performance Evaluation and Results

The performance evaluation for the proposed platform was done using Google lighthouse. The device used for performance check is an emulated desktop, with network throttling of 40ms TCP RTT, 10240 Kbps throughput (Simulated). User-agent (Host) is Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/89.0.4389.114 Safari/537.36 and the User-agent (Network) is Mozilla/5.0 (Macintosh; Intel Mac OS X 10_14_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/84.0.4143.7 Safari/537.36 Chrome-Lighthouse. The web interface received an overall performance rating of 92. The website was assessed for the following metrics: First contentful Paint, Time to interactive, Total Blocking time, Cumulative Layout Shifts, and Speed Index, Largest Contentful Paint and got *fast* rating under 4 out of 6 metrics and *moderate* rating for 2 out 6 metrics as shown in figure 3.

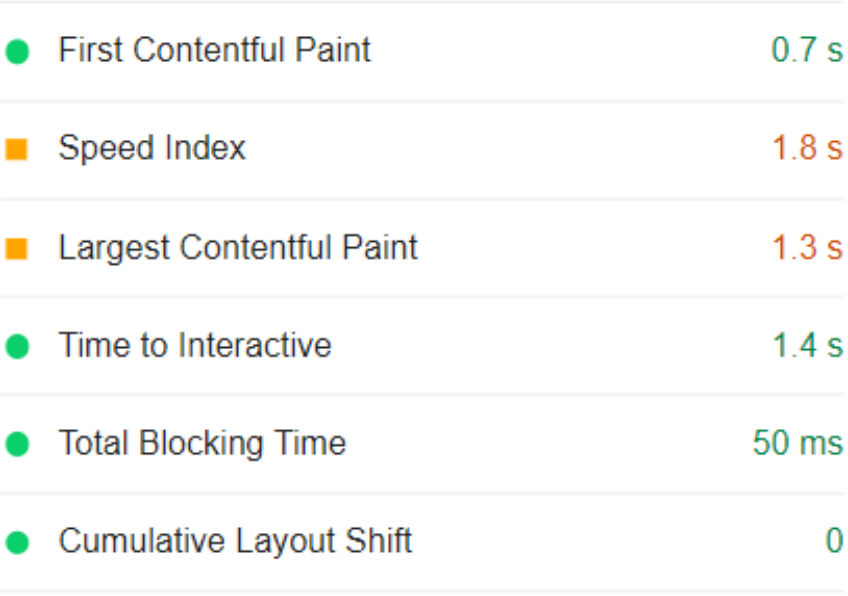


**Figure 3: Performance metrics for the web interface.**

In contest against the available alternatives, the proposed model provides greater set of features and proves to be more scalable and facilitates a great deal of automation as shown in Table 1.

| | **Blockchain** | **Record** | **Automated** | **Immutable** | **Scalable** |
|---|---|---|---|---|---|
| **SVM Model** | NA | Bank cheques | Yes | No | No |
| **King Abdulaziz University** | NA | Academic documents (any format) | Yes | No | Yes |
| **IDStack** | NA | Only PDF documents | Yes | No | Yes |
| **Hash and OCR** | NA | Physical Documents | No | No | No |
| **MIT DCC** | Bitcoin | Any | No | Yes | No |
| **ShoCard** | Bitcoin | Any | Yes | Yes | Yes |
| **EduCTX** | ARK | Student Credit Record | Yes | Yes | Yes |
| **Proposed Solution** | Ethereum | Any | Yes | Yes | Yes |

**Table 1: Comparative analysis with alternative solutions present in the industry.**

The platform built, offers a faster, more reliable, tamper-proof, and fraud-free solution to the document verification process of universities and employers. Once deployed, the data within the blockchain cannot be altered even by the university employees, therefore ensuring the authenticity of the displayed data. Moreover, the web interface has a straightforward and intuitive user interface, it offers single tap verification of data. The blockchain-powered architecture offers impeccable security and this, in turn, builds integrity and trust for the documents within the system. This solution outperforms manual

transmission of data and other alternatives, by all standards. Not only does it offer absolute security against counterfeit but also curtails wait-time by a very significant amount.

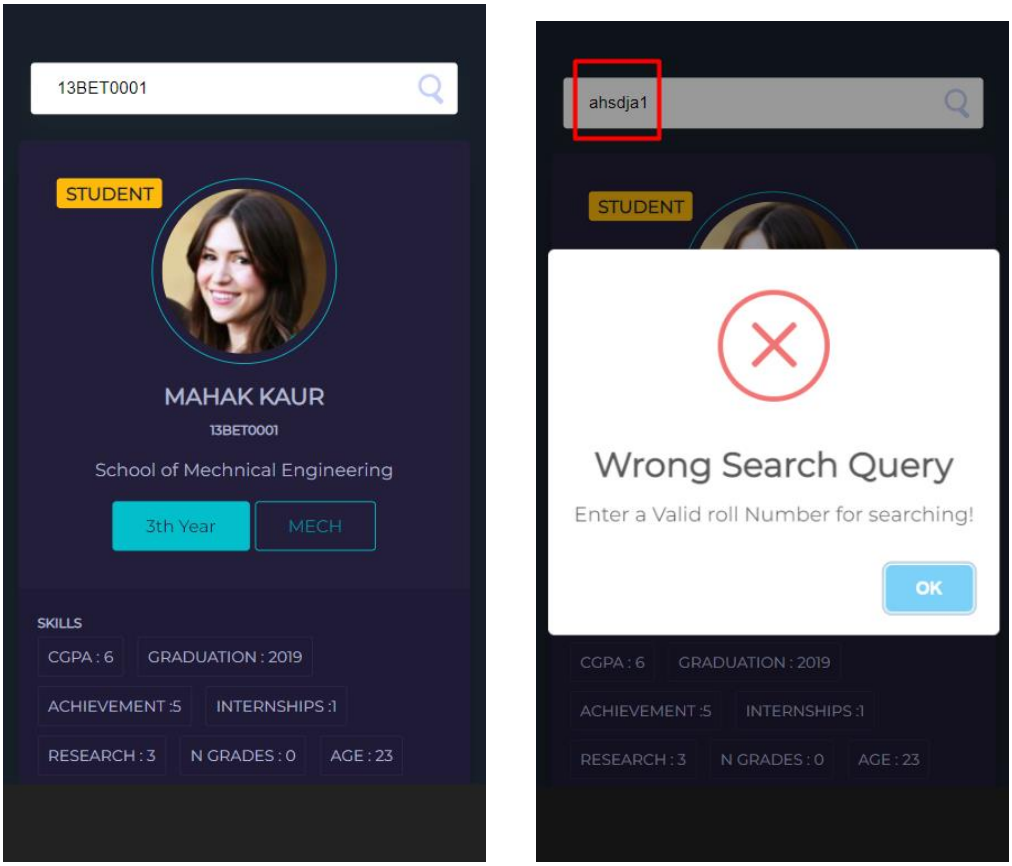


**Figure 4: The web interface displaying the student's information and detecting faulty queries.**

The web interface gives a thorough overview of a student's profile by giving information about the academic record, extracurricular record, industrial experience, research experience, and much more as shown in figure 4.

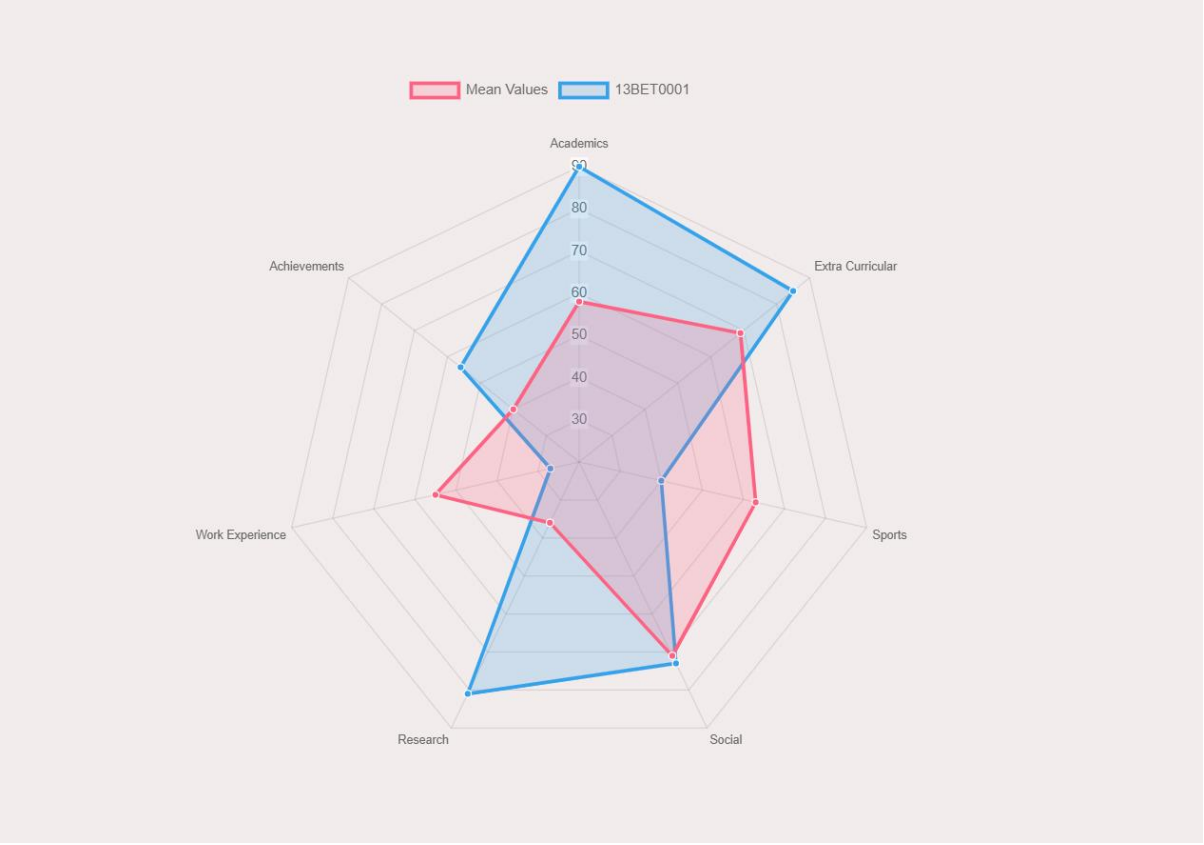


**Figure 5: Analytical overview of the student's performance.**

Additionally, the web interface offers, searching candidates based on their skill, which helps employers to shortlisted desirable students for their job role and more over provides an analytical overview of the student's abilities while comparing it to the average of all the students in the university/organization as shown in figure 5.

## 7. Conclusion and Future Work

This paper puts forward a platform that offers a secure and comprehensive alternative for employers to verify student data given by the university in a faster and more reliable way. The architecture is powered by an Ethereum blockchain and is offered as a web interface to the users build upon on javascript. The proposed platform has many verticals that could be exploited to produce a complete tool kit for academic institutions. The platform is being worked upon to offer, charts and graphs as an analytical overview of the student's performance and aptitude. Moreover, features such as offering to store the hash of the official academic transcript for verification and authentication purposes could be looked upon further. Moreover, the performance of the web interface can be further improved by using the industry best practices and removing certain unused libraries and styling in the code.

## 8. References

[1] Garain, U., & Halder, B. (2008, December). On automatic authenticity verification of printed security documents. In *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing* (pp. 706-713). IEEE.

[2] Brdesee, H. S. (2019). An Online Verification System of Students and Graduates Documents and Certificates: A Developed Strategy That Prevents Fraud Qualifications. *International Journal of Smart Education and Urban Society (IJSEUS)*, *10*(2), 1-18.

[3] Lakmal, C., Dangalla, S., Herath, C., Wickramarathna, C., Dias, G., & Fernando, S. (2017, September). IDStack—The common protocol for document verification built on digital signatures. In *2017 National Information Technology Conference (NITC)* (pp. 96-99). IEEE.

[4] Putro, P. A. W., & Luthfi, M. (2019, October). An authentic and secure printed document from forgery attack by combining perceptual hash and optical character recognition. In *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 157-162). IEEE.

[5] MIT Media Lab Learning Initiative and Learning Machine, "Digital Certificates Project," [Online]. Available: http://certificates.media.mit.edu/

[6] "ShoCard | Identity for a Mobile World", *Shocard.com*, 2017, [online] Available: https://shocard.com/.

[7] Turkanović, M., Hölbl, M., Košič, K., Heričko, M., & Kamišalić, A. (2018). EduCTX: A blockchain-based higher education credit platform. *IEEE access*, *6*, 5112-5127.

[8] Badr, A., Rafferty, L., Mahmoud, Q. H., Elgazzar, K., & Hung, P. C. (2019, June). A permissioned blockchain-based system for verification of academic records. In *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (pp. 1-5). IEEE.

[9] Nakamoto, S. (2019). *Bitcoin: A peer-to-peer electronic cash system*. Manubot.

[10] Chowdhury, M. J. M., Colman, A., Kabir, M. A., Han, J., & Sarda, P. (2018, August). Blockchain versus database: a critical analysis. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 1348-1353). IEEE.

[11] Buterin, V. (2017). Ethereum: a next generation smart contract and decentralized application platform (2013). *URL {http://ethereum. org/ethereum. html}*.

[12] Metcalfe, W. (2020). Ethereum, Smart Contracts, DApps. In *Blockchain and Crypt Currency* (pp. 77-93). Springer, Singapore.

[13] Wang, S., Ouyang, L., Yuan, Y., Ni, X., Han, X., & Wang, F. Y. (2019). Blockchain-enabled smart contracts: architecture, applications, and future trends. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *49*(11), 2266-2277.

[14] ConsenSys. (n.d.). *Metamask Wallet*. MetaMask Docs. https://docs.metamask.io/guide/.

[15] *Rinkeby Ethereum Test Network*. AirSwap Support. (n.d.). https://support.airswap.io/en/articles/2831385-what-is-rinkeby.

[16] Remix. (n.d.). *Remix - Ethereum IDE*. Remix. https://remix-ide.readthedocs.io/en/latest/.

[17] Wood, G. (n.d.). Solidity. https://docs.soliditylang.org/en/v0.8.3/.

[18] *web3.js - Ethereum JavaScript API*. web3.js - Ethereum JavaScript API - web3.js 1.0.0 documentation. (n.d.). https://web3js.readthedocs.io/en/v1.3.4/.

[19] Dannen, C. (2017). *Introducing Ethereum and solidity* (Vol. 318). Berkeley: Apress.

[20] Iyer, K., & Dannen, C. (2018). The ethereum development environment. In *Building games with ethereum smart contracts* (pp. 19-36). Apress, Berkeley, CA.

[21] Lee, W. M. (2019). Using the metamask chrome extension. In *Beginning Ethereum Smart Contracts Programming* (pp. 93-126). Apress, Berkeley, CA.

[22] Lee, W. M. (2019). Using the web3. js APIs. In *Beginning Ethereum Smart Contracts Programming* (pp. 169-198). Apress, Berkeley, CA.

[23] Zheng, G., Gao, L., Huang, L., & Guan, J. (2021). Application Binary Interface (ABI). In *Ethereum Smart Contract Development in Solidity* (pp. 139-158). Springer, Singapore.