EduChain:

Leveraging Blockchain, Federated Learning, and AI Fraud Detection for CBA

Presented By: Wanican Julian Okello

The 41st Annual Conference of the AEAA,
Skylight Hotel Addis Ababa, Ethiopia

Authors

Wanican Julian Okello
Software Engineer, UNEB

Dr. Oguti Ann Move
Lecturer, Soroti University

Opio Henry Joe
Data Warehouse Officer, UNEB

Presentation Roadmap

INTRODUCTION

- ❑ Overview
- ❑ Problem Statement
- ❑ Objectives

LITERATURE REVIEW

- ❑ Define Core Concepts
- ❑ Examine Related Works

METHODOLOGY

- ❑ Research Methodologies used

RESULTS

- ❑ Results achieved
- ❑ Implications

FUTURE WORKS

- ❑ Project Continuity
- ❑ Q&A

Project Background

2005-2019: Widespread adoption of the Competence Based Curriculum in Africa

Tertiary Institutions: Originally adopted by institutions of higher learning like universities, vocational schools (TVET) e.g. Nursing, healthcare, hospitality

Past 10 Years: CBC introduced in secondary schools across Africa. Rwanda 2015, Kenya 2017, Uganda 2020, Zambia 2013, Zimbabwe 2017

Implementation: Every country has mostly defined its own implementation framework as long as it adheres to the core principles of CBC

Who has adopted the Competence-Based Curriculum in Africa?

Countries Implementing CBA

Created with mapchart.net

Country	Level of Adoption	Key Focus Areas
Kenya	Primary, Secondary, Higher	National CBC rollout and CBA focus
Uganda	Primary, Secondary, Higher	CBC and CBA in schools and universities
Rwanda	Primary, Secondary, Higher	CBA for skill demonstration, aligned with CBC
South Africa	Secondary, Higher	OBE/CBA system, vocational and technical education
Tanzania	Primary, Secondary, Higher	CBC and CBA in schools and universities
Cameroon	Primary, Secondary, Higher	CBA in technical education
Senegal	Primary, Secondary, Higher	CBA and CBC in national curriculum
Zimbabwe	Primary, Secondary, Higher	Technical colleges, vocational programs
Malawi	Primary, Secondary, Higher	Vocational and technical education, CBA
Gabon	Primary, Secondary, Higher	Adoption in technical education
Egypt	Primary, Secondary, Higher	CBC and CBA for skill assessment
Mauritius	Primary, Secondary, Higher	CBA in schools and higher education
Nigeria	Secondary, Higher	CBA for vocational and technical education

Adopted From Various sources i.e. Ministries of Education of the various countries, publications on CBC implementation

CBA: What are the core principles of Competence Based Assessment?

EVIDENCE BASED ABILITIES

CONTEXTUAL ASSESSMENT

CONSTRUCTIVE FEEDBACK

CONTINOUS IMPROVEMENT

- Competencies should be clearly defined with knowledge, skills, and behaviors to achieve those competencies.

- Assessments should simulate real or near-real-world situations and tasks in practical contexts

- Learner abilities should be evidence-based and can be measured against various pre-defined criteria in every field.

- Learners should be individually involved and engaged in the assessment process, understanding their progress and taking ownership of the learning process.

- Teachers should always offer timely feedback that is constructive and actionable.
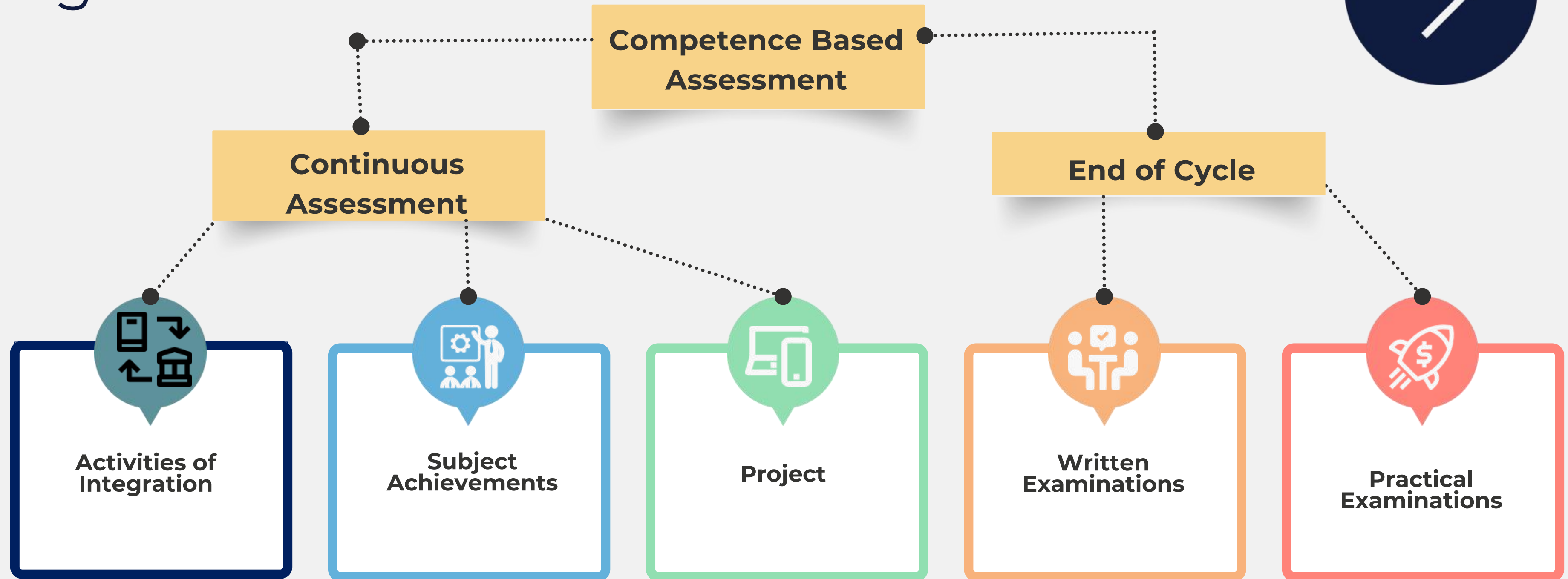
- The main goal of CBA is to encourage continuous improvement, and a timespan assessment of a learner should show progress over time

Competence Based Assessment is a derivative of the Competence Based Curriculum. It is a method of evaluating learners based on their demonstrated level of competence against a predetermined performance standard. (Idrissi, M. K., Hnida, M., & Bennani, S. (2016))

Excerpt: Implementation Framework of CBA in Uganda

Competence Based Assessment

Continuous Assessment

End of Cycle

Activities of Integration

Subject Achievements

Project

Written Examinations

Practical Examinations

IMPACT: CONTINOUS ASSESSMENT HAS TO BE FULLY "OUTSOURCED" TO TEACHERS AND EDUCATORS

Technological Implications to **National Assessment Bodies**

1.   Exponential Growth in Data Volume

The data collected for CA causes an exponential growth   in the overall assessment data volume due to the level of granularity required

2. Data Privacy Challenges Due to Decentralization

With the increased volume of data, and entrustment of teachers to the assessment process, there is a greater need to ensure data privacy for learners

3. Data Integrity Concerns

There is no guaranteed way to know that teachers will award valid scores to learners

4. Stakeholder Involvement

Learners, teachers, parents, and regulators need a platform to monitor performance in order to make timely interventions

Problem Statement

The introduction of Competence-Based Assessment has led to an exponential increase in the volumes of assessment data stored and managed by National assessment bodies, raised significant concerns in data privacy and integrity and challenged the need for consistent stakeholder involvement in the learners journey
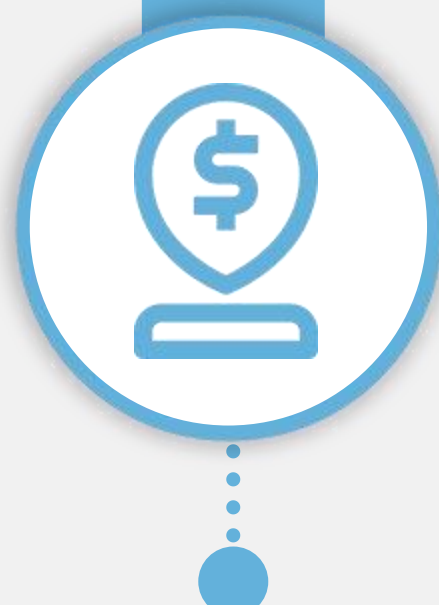
What are the objectives of EduChain?

01

BLOCKCHAIN

To achieve a highly secure peer approved transaction model in CBA database systems using HyperLedger Fabric Blockchain
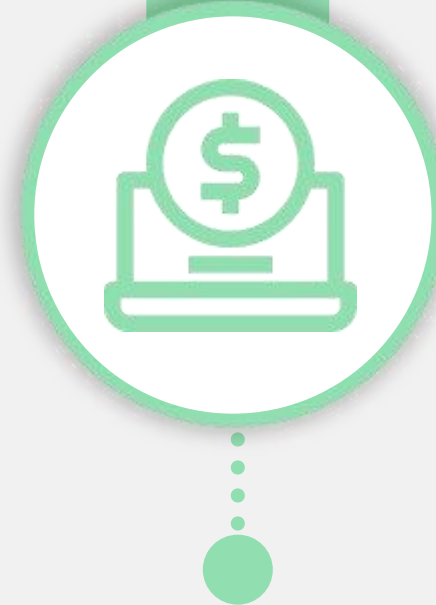
02

AI FRAUD DETECTION

To automate fraud detection on learner scores based on time series performance data and optimized AI models

03

FEDERATED LEARNING

To perform machine learning on de-identified data to improve model accuracy without compromising learner privacy

04

STAKEHOLDER INVOLVEMENT

Build an integrated multi-platform application accessible to learners, parents, teachers, schools and the national assessment bodies over mobile, web, desktop with a powerful backend engine

Implementations of **Blockchain** in the Education Sector

The most mature use case of Blockchain at the time of this presentation is credentialing; i.e., issuing and validation of student certificates.
These, however, are only in select American and European Universities and institutions. No publically available resource indicates such an undertaking in Africa.

EduChain will therefore be the first of its kind in an African institution

Student Data Management

Exam Integrity & Anti Fraud

Smart Contracts Tuition

Credential Issuing and Verification

A Blockchain refers to a <u>distributed</u>, <u>decentralized</u> and <u>immutable</u> ledger that records transactions in a <u>secure</u> and <u>transparent</u> manner

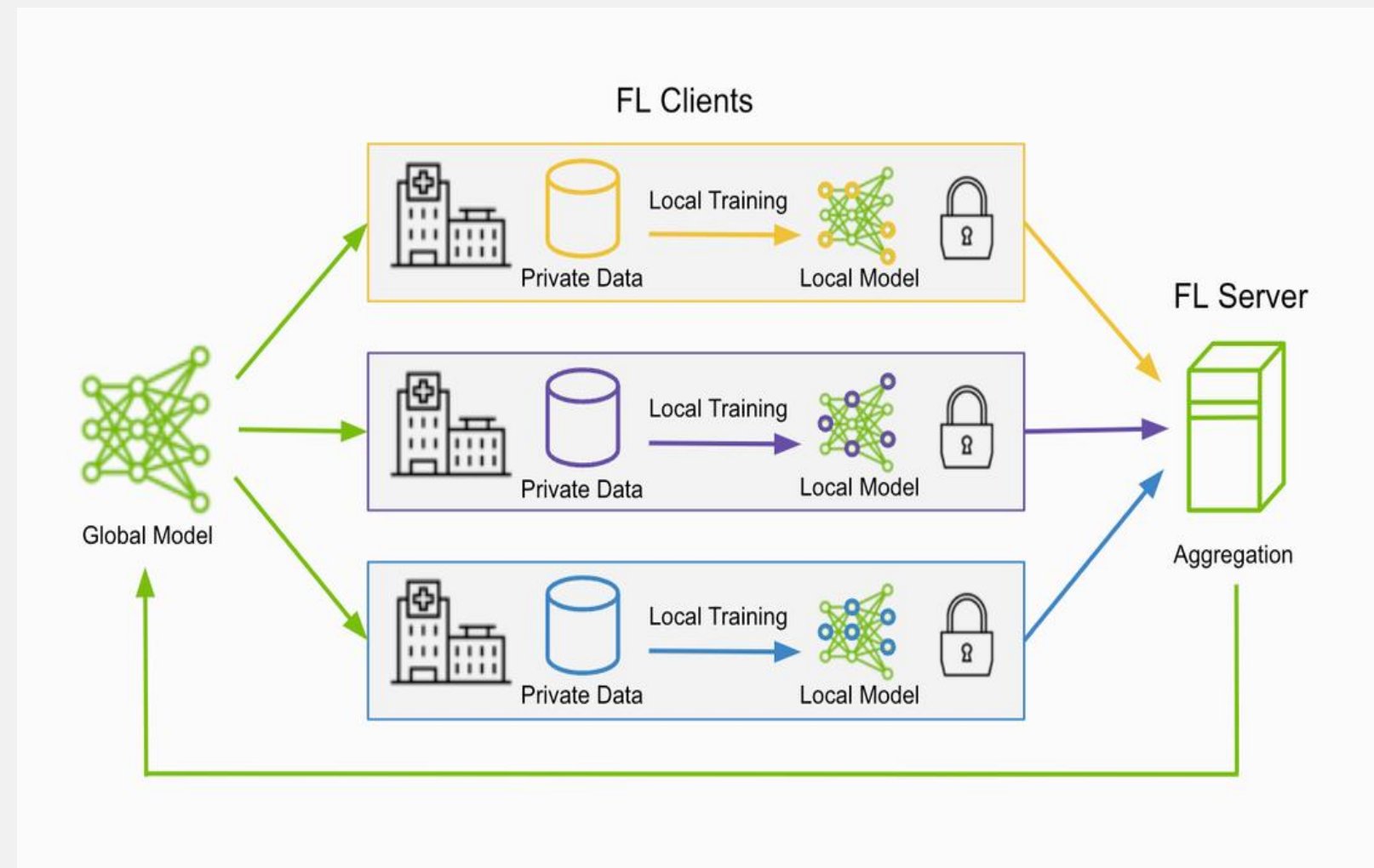# **Federated Learning** in the Education Sector

Since it's even newer than Blockchain, the only implementations are still in pilot stages all around the world

FL-MOOC: A version of Moodle that uses federated learning

EU H2020: Uses FL for risk prediction across schools

LEARNER-China: uses FL for distributed learning analytics

GRAINE: Japan Uses FL for distributed educational data analysis among schools

CHALLENGES
- Technologically Complex
- Synchronization of models between nodes can be expensive and network-intensive

AI **Fraud** Detection in the Education Sector

Most of the prominent online learning platforms use AI-based fraud detection tools

Remote Proctoring: ProctorU, Pearson VUE,
Plagiarism: Turnitin, Urkund, Copyleaks
Authorship: (LLMs) - Journals and Publishing houses

Limitations:
Specialized models work best in optimized scenarios

More data points = Higher Accuracy

Research Methodologies Used

1. Literature Review
   - ❏ Framework Guidelines from government organizations in charge of education
   - ❏ Documents from National Assessment Bodies
   - ❏ Documents from National Curriculum Development Organizations
   - ❏ Whitepapers on the proposed technologies
   - ❏ Publications on implemented technologies in EdTech in Africa (very scanty)

2. Experiments
   - ❏ Open source software tools and libraries

3. Case study of the Uganda National Examinations Board

Project Development Environment

To implement a proof of concept for this solution, we set up a virtual environment first on a local machine and then deployed it on a remote server for integration tests. The table below describes our development environment and tools used

TOOLS	DESCRIPTION/SPECIFICATION
LAPTOP COMPUTER	2.29 TB HDD 4GB GRAPHICS CARD 16GB RAM 11TH Gen Intel(R) Core(TM) i7-11800H @2.3GHz
AI FRAUD DETECTION & FEDERATED LEARNING	PyCharm 2022.3.1 Open Source Python Libraries SciKit, Keras, TensorFlow Federated (TFF), PyTorch KNN, Random Forests, Isolation Forests
BlockChain	Docker Container through WSL HyperLedger Fabric deployed on a local network in a docker container
DATA SET	Synthetically generated Dataset with 4 pre-determined falsifications i.e. Blatant Fraud, Imbalanced Fraud, Subtle Fraud, Copy Paste Fraud

Considerations

1. CBA Scores are descriptive and not explicitly numerical, therefore an initial data mapping has to be performed in order to transform the data to a more palatable format for consumption e.g.

| ACHIEVEMENT LEVEL	DESCRIPTOR	MAPPED DATA
Exceptional	Demonstrates an **extraordinary** level of competency by applying innovatively and creatively the acquired knowledge and skills in real life situations.	5
Outstanding	Demonstrates a **high** level of competency by applying the acquired knowledge and skills in real life situations	4
Satisfactory	Demonstrates an **adequate** level of competency by applying the acquired knowledge and skills in real life situations.	3
Basic	Demonstrates a **minimum** level of competency in applying the acquired knowledge and skills in real life situations.	2
Elementary	Demonstrates below the **basic** level of competency in applying the acquired knowledge and skills in real life situations.	1

Experiments For AI Fraud Detection

On a dataset with 300,000 candidates with nine subjects per candidate, we achieved the following performance on the three **AI Fraud Detection** Models tested;

Feature Engineering: We also used candidate index numbers and sitting order to add a data point and increase the accuracy of the model

Data Points: Candidate Unique Identifier, Scores per assessment per submission, seating position
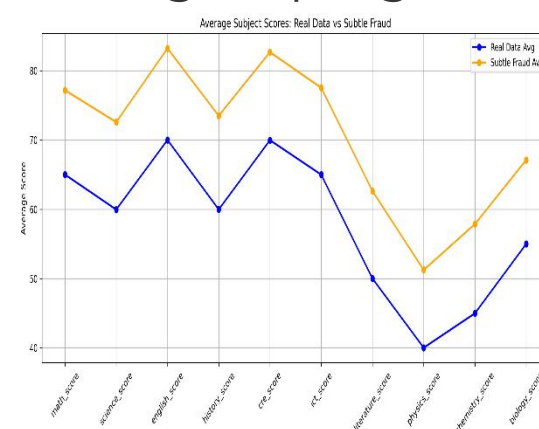
Copy Paste Fraud

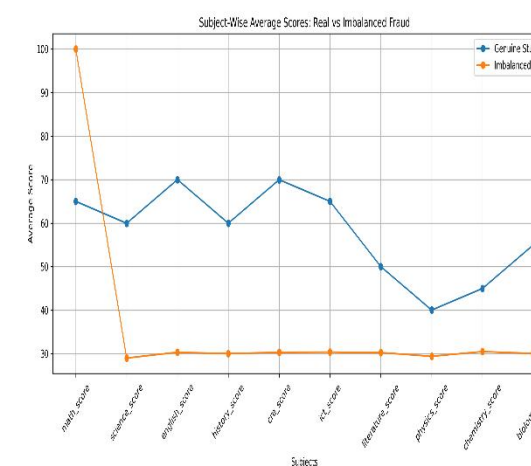Almost no variance between all the data points especially after grouping
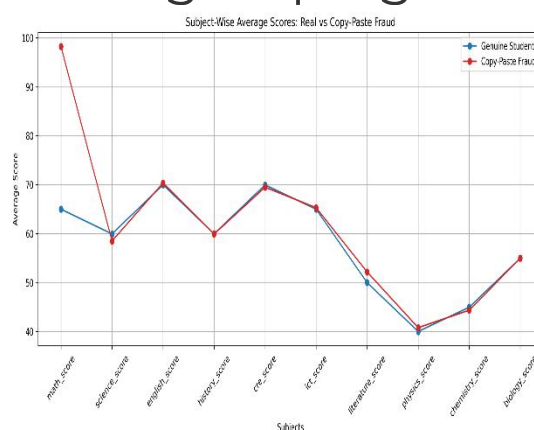
Subtle Fraud

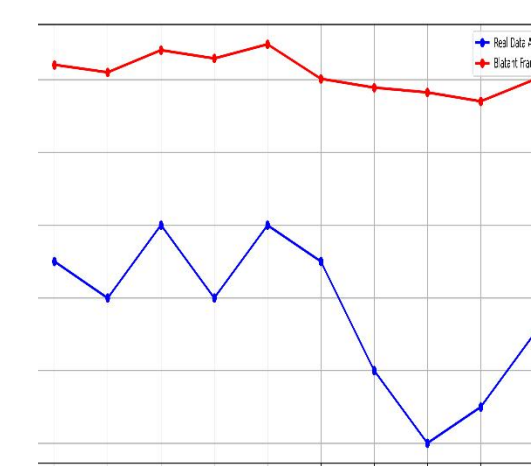Very Low variance between all the data points especially after grouping

Imbalanced Fraud

Sudden deviation from the normal distribution

Blatant Fraud

Glaring deviation from the normal distribution

Results For AI Fraud Detection

On a dataset with 300,000 candidates, nine subjects per candidate and one data point per subject, we achieved the following performance on the three **AI Fraud Detection** Models tested;

FRAUD TYPE	ISOLATION FOREST			KNN MODEL			RANDOM FOREST MODEL		
	PRECISION	RECALL	F1 SCORE	PRECISION	RECALL	F1 SCORE	PRECISION	RECALL	F1 SCORE
Copy_Paste Fraud	1.0	0.039	0.076	1.0	0.12	0.214	1.0	1.0	1.0
Subtle Fraud	1.0	0.448	0.619	1.0	0.204	0.339	1.0	0.872	0.932
Imbalanced Fraud	1.0	1.0	1.0	1.0	0.116	0.208	1.0	1.0	1.0
Blatant Fraud	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0

Table showing Performance of the Isolation Forest Model, KNN Model and the Random Forest Model

PRECISION: Proportion of True Positives

RECALL: Proportion of True Positives with respect to all the positives

F1 SCORE: Harmonic mean of the Precision and Recall

DISCLAIMER:
AI is not a foolproof method to detect fraudulent scores, however, it is a tool to efficiently flag potentially fraudulent scores

Experiments for Data Integrity Using HyperLedger Fabric

Simulated an environment with 4 organizations (Data Resting Points), each organization with 5peers. We maintain the total number of records at 300,000. Our aim was to test the fabric network performance and security

Performance Metrics on Simulation Environment and Server Environment

Metric	Simulation Environment	Server Environment (64GB RAM, GPU, SSD)
Throughput	150 - 250 tx/sec	700 - 1100 tx/sec
Latency per Transaction	450 - 700 ms	100 - 180 ms
CPU Usage	~85% of 4 CPUs	~60% of 16+ CPUs
RAM	8 - 10 GB	25 - 35 GB
Disk I/O	Moderate to High	High
Network I/O	High	Very High
Total Processing Time	~20 - 33 minutes	~5 - 7 minutes

Security Tests on the Fabric Network

Test #	SECURITY TEST NAME	Expected Behavior	PASS/FAIL
1	Identity & Certificate Validation	Reject invalid, expired, unauthorized certificated	Passed
2	Endorsement Policy Bypass	Transaction rejected if it lacks required endorsements	Passed
3	Endorsement Policy Bypass	Transaction rejected if it lacks required endorsements	Passed
4	ChainCode injection and Integrity	Only authorized users can deploy/upgrade ChainCode	Passed
5	Private Data Access Control	Unauthorized orgs can't read private data	Passed

Conclusion: What have we achieved so far with respect to our objectives?

1. Proof that AI models can detect 4 different types of CBA scores fraud

    Copy Paste Fraud to 7.6%(IF) – 100%(RF),

    Subtle Fraud to 33% (KNN) - 93%(RF),

    Imbalanced Fraud 20%(KNN) – 100%(RF),

    Blatant Fraud 0%(KNN) - 100%(RF)

2. It may be expensive to run a full scale Blockchain network on all CBA scores but it is feasible and secure to run on a subset of the data e.g. the final scores at the end of an education level

How far are we in the larger scheme of the project?

Literature Review

Study Related Works and build upon it

Stakeholder Involvement

Bring core stakeholders on board

Pilot
Pilot EduChain in a few schools to improve it

Project Origination

Study the problem and set objectives.

Technical Feasibility Tests

Test feasibility of technologies in proposed environment

Build EduChain

Build the overlying application and integrate underlying technologies and existing school applications

Deploy

Deploy EduChain to production

Future Direction of **EduChain**

"Every great innovation was once an idea..."

> EduChain is a bold, daring, and ambitious idea that aims to change the narrative of the use of technology in Africa specifically in Education. We have to move beyond the role of playing catch-up with the rest of the world, and being at the forefront of solving real problems!

Proof Of Concept – **DONE!**

NEXT STEP?
Build at scale and Pilot EduChain in a live environment

Thank You For Listening
Let's Discuss!

@256 701 077145 - julian.okello@uneb.ac.ug