Fraud Management System For UID Aadhaar

Pragati Mynampati, Dr Girijamma H A

Abstract—UID Aadhar is a 12-digit unique identity number assigned to individual residents of India based on their biometric and demographic data submitted at the citizens enrollment centers, India. The current Aadhaar system issues the UID number after verifying the de-duplication of demographic and biometric data of the citizens in the system. There are many evidences that we have seen such as, spurious documents of proof of Identity, proof of address and proof of residences are being generated by the illegal immigrants, terrorists, malicious residents in order to meet their deceitful ends and appearing in the enrollment process and getting a valid UID Aadhaar number. This paper discusses broadly various enrollment frauds that have been observed in the past and present times. It discusses how such frauds can be eliminated by introducing a fool proof fraud management system in the current Aadhar system. It also discusses that various AI & ML techniques that can be applied in detection of such enrollment frauds.

Index Terms— Fraud Management System, UIDAI, Aadhaar, UID enrolment, Machine Learning, Deep Learning, Supervisory techniques, Unsupervised techniques, ROC-AUC score, Confusion matrix, F1 score.

1. Overview

Nowadays, due to the meteoric increase in the population, it is necessary to identify each individual uniquely. The ability to identify a person uniquely, reduces the chances of getting deceived by malicious residents and increases the chances to avail various social benefits and security. It is necessary to protect one's own identity to avert crimes caused due to identity theft.

To avail any services provided by the government, one should prove that, they are the genuine individual they claim to be. This can be done with the help of UID Aadhaar card, which provides the proof regarding the person's genuine identity.

There are several advantages of having an Aadhaar card such as,

- Universal address proof: Address proof is required by various organizations. Aadhar card can be used in such situations.
- Can open a bank account, invest in mutual funds or buy an insurance policy.
- Receive government benefits directly: with the help of Aadhaar card, this process is simplified by accessing these benefits directly by linking Aadhaar card to the bank account.
- Digital locker: Digi-locker system has been initiated by the government, so that the individuals can store their personal documents. This can be done by linking the Aadhar card and storing the documents in the government's secured server.

- *Pragati Mynampati is currently pursuing bachelor's degree program in computer science in RNSIT,VTU (Visvesvaraya Technological University) ,India, PH-(91)9740463355. E-mail: shammysham77@gmail.com.*
- *Dr. Girijamma H A is currently a professor in computer science department in RNSIT, VTU, India, PH-9480031494. E-mail: girijakasal@gmail.com.*

2. Current UID Enrollment System

Several UID enrolment system consists of the enrolment client, enrollment processing server, backend databases, and big data servers for storing large volumes of demographic and biographic information of the total population of the country

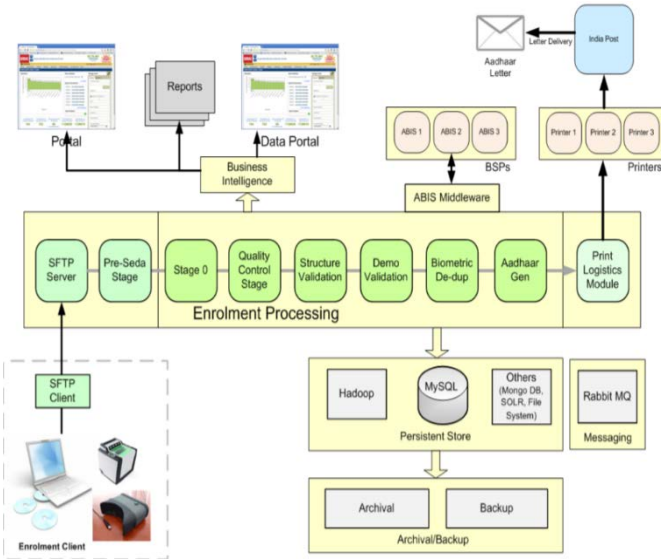

Figure 1: UID Enrollment System (Reference 2)

2.1 Enrollment Client – The Enrolment system is based on the multi-registrar model. This means that, the software "Enrolment Client" is created by the UIDAI and given to the appointed registrars to be used at their "Enrolment

Agencies". The Enrolment Client is a software that is used for first time enrolment and to update the data. The Enrolment Client does all the necessary tasks such as,
Recording resident's demographic details: records the resident's name, address etc.
Biometric data capture: this captures the biometrics of the residents. It includes capturing of 10 fingerprints, both irises and a photography of the face.
Biometric data quality check.
Resident data validations: This ensures that all the demographic and biometric data entry is done in the presence of the resident. The residents can cross verify the data before it is entered into the database.
Corrections and Updates: This allows manipulation of data, even after the enrolment process is done to correct the erroneous data.
The enrolment client software also provides security features such as data encryption, data storage and finally exports the data packets to centralized enrolment server.

2.2 Enrollment Server – The These servers manage the enrolment packets, validates the data and does quality checks. It also de-duplicates data using multiple ABIS (Automated Biometric Identification System) to make sure there is no redundancy of biometrics involved, in case the citizen enrolls for second time from other location of citizen enrolment center.
Aadhaar enrolment servers can handle more than 1.5 million enrolments everyday and are built to manage data stores in hundreds of terabytes.

2.3 Uploading Packets – The Enrolment client generates packets that contains the biometric and demographic data. These packets are encrypted using the 2048-bit PKI (Public Key Infrastructure). Each packet is around 3-5 MB in size. Packets are then uploaded to UIDAI data centers for further processing.

These packets are scanned for various malwares and are tracked to see if they are from trusted enrolment client centers. Any packet that looks apocryphal is rejected. The genuine packets are then sent into the production zone of data center. The valid packets are decrypted and are sent to the Enrolment processing workflow.

2.4 Packet Archival – All the packets that are sent by the Enrolment Client should be stored in the data centers.
The packets are always stored in the encrypted format even within the archive. Once the Aadhaar number is issued, there's no need for accessing the raw packet again. Therefore, the original enrolment packet in its encrypted format is archived and not accessed until any manual inspection or investigation occurs.

2.5 Data Validation –After the packets are scanned for malwares and are verified as genuine, validation of data inside the packet is carried out to ensure the authenticity and correctness. Some of the validations that are carried out are,

- Demographic data validations: It includes phonetic matching of name, valid age range check, address structure check against postal pin code and geographic boundary master data.
- Biometric data quality checks: Biometric deduplication checks are performed with Iris, finger prints and face photos captured during enrollment.

2.6 Biometric Deduplication – The Aadhaar card is issued to persons whoever submits unique set of 10 fingerprints, 2 IRIS prints and a photograph of the face during enrolment. The Enrolment system consists of three independent ABIS (Automated Biometric Identification Systems) from three different vendors are used to ensure higher levels of accuracy and performance. The ABIS system compares the resident's biometrics with all existing biometrics in their gallery to find duplicates, If any

2.7 Aadhaar Number Generation –The first 11 digits generated are random numbers and the last digit is generated using checksum based on Verhoeff algorithm. This ensures the uniqueness of the number generated and makes sure that no resident gets the same number. The Permanent identification number is assigned to the enrolled citizen only after successful biometric de-duplication.

3. Enrollment Fraud Scenarios

The objective of fraud management system will be to ensure that fraud enrolment is detected and prevented.
A fraud management solution is required to detect and reduce enrollment related frauds.
The fraud management solution should be able to detect frauds such as:

3.1 Misrepresentation of information:

Illegal immigrants, terrorists who enter into the country will enroll into the UID Aadhaar by producing fake supportive proof of identity, proof of residence documents. The enrolment system cannot identify these intentional misrepresentation of demographic information (unless there is similar de-duplication of data). These illegal immigrants, by producing fake original documents at CEC (Citizens Enrolments Centre) or by bribing the CEC agents or by stealing the credentials of CEC agents, these malicious residents get enrolled in the program. This means that, they cannot be identified by bio-metric de-duplication checks in the enrollment server.

3.2 Enrolments for non-existent individuals:

The illegal immigrants can collect the dead citizen's demographic details and get enrolled to the UIDAI

system, if the dead individual didn't get enrolled into the UIDAI system yet. If the dead individual has already enrolled to the UIDAI system, then enrolment system biometric de-duplication ABIS module detects and rejects the application.

3.4 Enrolments outside the country:

This is strange enrollment scenario where the fraud racket team with the support of the respective country's CEC agents, getting their credentials either by bribing or stealing, carry out mass enrollment activities outside the country for the illegal immigrants. Such frauds can be detected by tracking the IP address of the enrollment client. So, we need to capture IP address of the desktop enrollment client machine into our proposed data set.

3.5 Enrolment during unusual hours of the day:

This is additional time stamp parameter which we need to add into data set. All the CEC's enrollment activities must be closed after a time window specified by the government.

4. Proposed Fraud Management System

The architecture of the proposed fraud management system is shown below. It mainly consists of three components. First,
Big Data platform used for enrollment. Second, Fraud Management System platform. Third, dash boards and reporting platform.

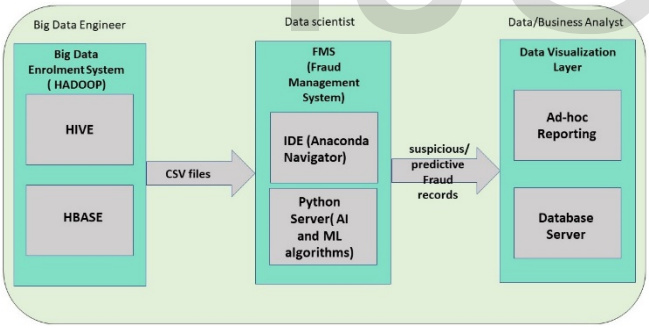

Figure 2 : FMS architecture

Fraud related models are built with the help of python-based machine learning and deep learning algorithms. The source data for the enrolment fraud is the Hadoop database in which the structured (demographic details) stored in HIVE and unstructured data repositories (biometric data) stored in HBASE is used. At the end of business hours of each day, the enrolment data along with Aadhar number is pulled through CSV files and fed into fraud management system (here after, FMS). A separate RDBMS is maintained in the FMS system to store every day's enrolled data received through CSV files. Fraud prediction / detection models are applied on the entire RDBMS. The fraud management system identifies the potential records which are suspected to be fraud and are further moved into data visualization layer for storing purpose useful for further analysis and viewing.

5. Preparation of Data Collection for FMS

Enrollment ID	Full name	Gender	Age	DOB	is DOB declared/ verified ?	DOB proof	Address	e-mail	Mobile #	PinCode	Father/ Mother/ wife/ husband	F/M/G/H Name	F/M/G/H Adhar #/ EID #
234565436543	Mary Steven	Female	29	4/5/90	yes	birth certific	#705,oak ap	mary@g	9090097767	560032	father	John Steve	678987678909
143256788765	Rose Heigle	Female	30	3/2/89	yes	birth certific	flat no.12,o	rose@gn	9887765658	460098	mother	Kate Grah	546789876564
453665477654	Alex Stamm	Male	28	3/4/91	no	no	door no 23(	kate@gn	9969353432	780034	wife	Lily Stamn	889754345432
765487678976	Sam Nichola	Male	25	2/6/94	yes	birth certific	#210,b bloc	sam@gn	9879876574	230089	father	Andy Nich	897687675675
657534543234	Ruby Evens	Female	35	8/5/84	yes	birth certific	flat no.45,p	ruby@gr	8667546324	670098	husband	Jared Ever	987876765453

The above excel data represents the normal enrolment system data column features in the Aadhaar form that won't detect the malicious residents who entered into the enrolment system. Therefore, we need more attributes to capture and feed into the FMS system, to detect the fraud enrolments.
Some of the attributes (shown in red) that can be augmented to the normal enrolment system data which can help us capture anomalies in the enrolment shown below.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	342343	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no

5.1 Case 1: POI (Proof Of Identity) Category

An individual proves that, they are the genuine individual they claim to be by, giving the proof of identity, but this can be misrepresented by producing fake POI by different means such as, a fake Gazetted officer certified POI document will be submitted during Aadhaar enrolment. This category of POI information should be entered into the enrolment database and should be passed into the FMS.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	342343	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no

Here, Alex with no POI proof is more susceptible to be a fraud. He has taken the advantage of lack of Authentic

proof of records trying to get alternative evidences from the Gazetted officer. Here, the Gazetted officer can be either bribed or deceived by showcasing the fake evidences. Alex with the help of Gazetted officer signed certificate, gets enrolled to the Aadhaar system.

5.2 Case 2: POA (Proof Of Address) Category

This is required to prove that the address submitted by the resident is genuine and not a fake one. In some cases, the illegal immigrants try to obtain the POA document by opening an account in a private bank, with the help of an introducer who already has an account in the bank. This can be used as a POA during enrolment process, which is a fake one.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	342343	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no

5.3 Case 3: Enrolment Time Stamp

This field records the timing when the enrolment of the applicant was done.

In general, the enrolment process takes place at citizen enrolment center from 10:00 AM to 5:30PM. Any enrolment which is taking place beyond the registration time is considered as a fraud practice.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	342343	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no

Here, Rose registered at 8:40 PM which is after the registration time window. Hence, it can be deduced that this is a fraud practice.

5.4 Case 4: Enrolment client machine IP address

This ensures that the enrolment is done in a genuine enrolment client system inside the country. Enrolments that are done outside the country has different IP address and can be classified as fraud.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	342343	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no

Here, Sam registered in an Enrolment Center, where the system's IP address contradicts the International network IP address assignations.

5.5 Case 5: Enrolment Agent Code

The code of the enrolment agent in the citizen enrolment center should be passed to the FMS system. This helps us in identifying if the enrolment agent is genuine or not.

For example, many of the fraud enrolments takes place by the illegal immigrants by bribing the enrolment authorities and agents. There are some cases found in Aadhar system where, the biometric credentials (fingerprints) of enrolment agents can be bribed or stolen and are used in fake enrolment centers, running at internet cyber centers.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	432654	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no

Here, Rose and Alex enrolled under the same enrolment agent. Since, we know that Rose is already involved in fraud enrolment (Enrolment time stamp) with the help of enrolment agent. So there are chances that Alex who has registered under same enrolment agent maybe involved in fraud enrolment too, with the help of enrolment agent who might be a fraud as well.

5.6 Case 6: Introducer's EID

This field also verifies if the introducer is genuine or not.

In some cases, The illegal immigrant who has already got his/her Aadhar card issued can act as introducer to another illegal immigrant.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	432654	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no
Olivia Sui	968362529328	voter ID	voter ID	2:09 PM	13.32.44.045	983893	no	878679875643	no

Here, Rose and Olivia have the same Introducers. Since Rose is already a fraud, chances are there that Olivia is a fraud too.

5.7 Case 7: Based on Gazetted officer's certificate

Some residents who do not have certain proofs (identity, DOB, address), take a certificate from the Gazetted officer, claiming that they are the genuine citizens. The officer verifies the details of the applicant and gives the certificate to them.

In some cases, The Gazetted officer, can be deceived by the fake certificates given by the residents, to claim the identity certificate.

Full name	Aadhaar #	POI proof category	POA proof category	Enrollment Time stamp	Enrollment M/C ip Address	Enrollment Agent Code	Physically challenged	Introducer/HOF Enrollment #/Adhar #	Gazeted Officer Cer.based
Mary Stevens	329876567384	passport	passport	10:30 AM	13.32.44.125	899667	no	567876543657	no
Rose Heigle	289387645378	passbook	passbook	8:40 PM	13.32.44.235	432654	no	878679875643	no
Alex Stamm	997867545632	no	no	11:23 AM	13.32.44.185	432654	no	598709735642	yes
Sam Nicholas	238753497235	passport	passport	3:07 PM	27.96.95.155	988767	no	987654345353	no
Ruby Evens	537265398457	passport	passport	4:08 PM	13.32.44.165	986437	no	897687698767	no
Olivia Sui	968362529328	voter ID	voter ID	2:09 PM	13.32.44.045	983893	no	878679875643	no

Since the Gazetted officer can be deceived or either bribed, it is necessary to check the documents submitted by the resident to the officer, is genuine or not.

6. Machine Learning and Deep Learning Fraud Detection Techniques

The AI & ML techniques can be broadly classified into two types. One being Supervised and other being Unsupervised.

6.1 Supervised Techniques: This technique is applied when sample data regarding fraud transactions records are available and are identified and clearly labelled as fraud enrolments along with normal enrolment data.

6.2 Unsupervised Techniques: This technique is used when we have no information of the fraud transactions when the data set is given. We need to predict the fraud records either by classification or by clustering. Prediction accuracies are poor with un-supervisory algorithms compared to supervised techniques.

In both models,

- Total data is segregated and separated as training data and test data with a ratio of 80%:20%.
- Both training and test data contains mix of samples or fraud and normal records (supervised learning)
- Training set is used to build the ML model.
- The test set is used to validate and predict the performance of the model.
- Since fraud is rare anomaly, we need to work with highly imbalanced data sets

As the fraud records are few in number when compared with genuine enrollment records, the data set applied to the system is highly imbalanced records for which we have identified few proven anomaly detection techniques such as Random Forest, XGB Classifier.

Isolation forest and One class SVM techniques are more suitable for detection of such rare frauds. We also suggest some of the deep learning algorithms such as back propagation neural networks for supervisory technique and auto encoder algorithms for un supervisory technique can be applied

7. Summary and Conclusions

The above said ML algorithms are applied and prediction accuracies are compared against each and every algorithm suggested. The performance metrics like confusion matrix, accuracy, precision, F1 score, ROC-AUC scores are captured for each algorithm and based on these figures the best algorithm can be chosen for enrolment fraud detection. In the beginning of the training models, as we don't have any clue on fraud enrollments un-supervisory techniques are applied. With the results of predicted suspicious fraud records after verifying in the back-ground checks with the individuals in the reality, many of the frauds are ascertained and are labelled. After having sufficient labelled fraud transaction records, the system can be trained with supervisory techniques and more accurate predictable fraud transactions can be obtained. As the enrollments for identification of citizens keeps increasing in the system, both supervisory and un-supervisory algorithms are running in parallel. This is an iteration process that keeps executing throughout the enrollment period.

8. References

[1] "Aadhaar technology and architecture: Principles, Design, practices and key lessons" – March 2014, UIDAI planning commission, Government of India.

[2] "Aadhar technology and architecture", March 2014, page 43, Figure 3: enrolment module overview.

[3] UIDAI https://uidai.gov.in/my-aadhaar/about-your-aadhaar/aadhaar-generation.html

[4] "Introduction To Aadhaar, Digital Identifiers" by Ashok Kumar.

[5] Modular Open Source Identity Platform - https://github.com/mosip

[6] Aadhaar-Wikipedia: https://en.wikipedia.org/wiki/Aadhaar

[7] Architecting World's Largest Biometric Identity System - Aadhaar Experience by Dr. Pramod Verma.