Big Data Applications in Aadhar Card Fraud Detection

K. Ramya[1*], A. Sumathi[2]

[1,2]Dept. of computer science Navarasam Arts and Science College for Women, Arachalur, Tamil Nadu, India

Corresponding author: ramyarasu12@gmail.com

Abstract- Big Data is playing a very significant role to take any industry forward. In the fraud detection, automated fraud detection tries to collect all information to reduce in aadhar card frauds by doing analysis and data mining of Big Data. This paper investigates the benefits of Big Data technology and main methods of analysis that can be applied to the particular case of fraud detection in aadhar card. This paper hereby addresses aadhar card fraud detection via the use of data-mining techniques in classification of, Naive Bayesian (NB), c4.5, and Back Propagation (BP) analyze the customer data. In order to identify the patterns that can lead to frauds. Upon identify all sectors, adding a Aadhaar card include Name, Age, Date of Birth, Aadhaar Number ,Gender ,Photograph ,Residential Address, that are stored in data base according to biometric data are Fingerprints and Iris scan. Representing the Aadhaar number Details stored in the database in Fingerprints, Iris scan. Finally Aadhar card frauds are identified and detected using data mining algorithms.

Keyword: Big Data Analytics, Big Data Applications, and Aadhar card fraud detection, Classification Algorithm

I. INTRODUCTION

A. Big Data Analytics
Big Data is very familiar term that describes voluminous amount of data that is structural, semi-structural and sub structural data that has potential to be mined for information. The problem starts during data acquisition, when the large amount of data require us to make decisions about what data to keep, what to discard and how to store, so that data can be kept reliable and accurate. Big data analytics is a technology that searches useful information such as a relation rule, a hidden value from huge data. When data volumes reach big data proportions, parsing it for meaningful information requires very powerful data analytics.
The analytics tools perform the following activities:
- Collects data from many enterprise sources.
- Performs a deeper analytics on the data.
- Provides a fine view of security information.
- Achieves real-time analysis of streaming data

B. Big Data Applications
Big data has found many applications in various fields today. The major fields where big data is being used are as follows.

In Government
Big data analytics has proven to be very useful in the government sector. Big data analysis played a large role in Barack Osama's successful 2012 re-election campaign. The Indian Government utilizes numerous techniques to ascertain how the Indian electorate is responding to government action, as well as ideas for policy augmentations real-time analysis of streaming data.

In Banking
The use of customer data invariably raises privacy issues. The uncovering hidden connections between seemingly unrelated pieces of data, big data analytics could potentially reveal sensitive personal information. Research indicates that 62% of bankers are cautions in their use of big data due to privacy issues.

In Stock
A private stock exchange in Asia uses in database analytics to establish a comprehensive system to detect abusive trading patterns to detect fraud.

In Fraud detection
Fraud detection is one of the most compelling big data applications, big data platforms that can analyze claims and transactions in real time, identifying large scale patterns across many transactions or detecting anomalous behavior from an individual user, can change the fraud detection game.

In Health care
Traditionally, the health care industry has lagged behind other industries in the use of big data, part of the problem stems from resistance to change providers are accustomed to making treatment decisions independently, using their own

clinical judgment, rather than relying on protocols based on big data.

In Marketing
Marketers have begun to use facial recognition software to learn how well their advertising succeeds or fails at stimulating interest in their products. A recent study published in the Harvard Business Review looked at what kinds of advertisements compelled viewers to continue watching and what turned viewers off. Among their tools was "a system that analyses facial expressions to reveal what viewers are feeling."

II. AADHAR CARD FRAUD DETECTION

The aadhar card contains a number of details itself. Some details are mentioned on the card where as others stored in aadhar's database. Various authentication request types of such as demographic and biometric details. Demographic data includes name, age, and date of birth, gender, and address of the resident, phone number, and email_id. Biometric data includes photograph of face, finger print, and iris scan.

Aadhar card applied for bank accounts, driving license, insurance policies, railway reservation, and applying PAN card, etc., most frauds happen because people using aadhar don't have enough clarity on how much information to share. Every year the number of fraudulent credit card, bank account, and fraudsters also use aadhar mobile phone linking to create fake aadhar card.

C. De-Duplication of Aadhar
Deduplication of aadhar contains two types of data stored in data base such as,

- Demographic De-duplication
- Biometric De-duplication

Demographic de-duplication
Demographic De-duplication is used to identity, trivial duplicates or cases of duplicates arising from error or ignorance. It used for children under the age of five years, as biometric data is not captured for them.

Biometric de-duplication
Biometric De-duplication is the primary method of identifying duplicates. It improves data accuracy. Biometric details are not match the aadhar card, to identifying fake biometric De-duplications.

III. CLASSIFICATION ALGORITHM

Classification is perhaps the most familiar and most popular data mining technique. Example of classification applications include image and pattern recognitions, detecting faults, loan approval in industry applications. Classification as a mapping from the database to the set of

classes, the classes are predefined, are non overlapping and partition the entire data base.

In this paper using the classification algorithms are: naïve Bayesian (NB), C4.5 and back propagation algorithms used.

D. Naive Bayesian Classification
The contribution by all attributes is independent and that each contributes equally to the classification problem, a simple classification scheme called naïve bayes classification. Bayes approach is easy to use and handle missing values by simply omitting that probability when calculating the membership in each class.
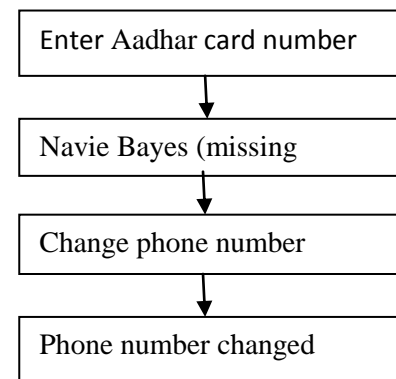

Figure: 1. Naive bayes classification

In the demographic data include name, age, date of birth, gender etc., in the Bayesian classification algorithm used to missing values, and the independent attributes are declared it. Once the aadhar card demographic details in anyone change or update the aadhar database focus Bayesian classification, phone number or address change to use details stored in aadhar data base.

E. C4.5
The decision tree algorithm based on C4.5. C.4.5 can help not only to make accurate perditions' from the data but also to explain the patterns in it. It deals with the problems are accurate prediction, missing data, pruning and generating the rule.

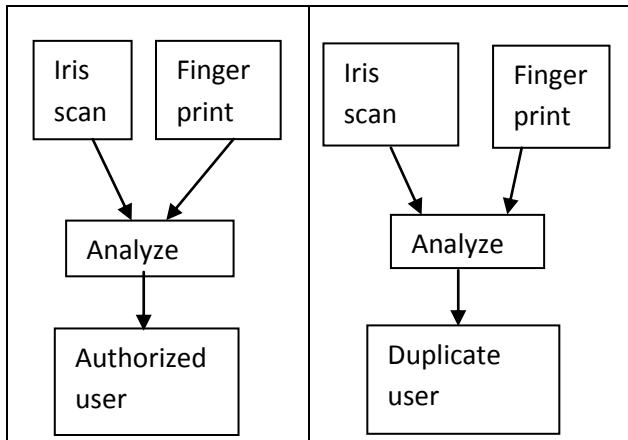| Original aadhar card	Duplicate aadhar card

Figure:2 Find out duplicate user using C4.5 algorithm

C4.5 allows classification and Aadhaar card in biometric details are photograph, finger print and iris scan. Once the finger print and iris scan scanned and stored the database that is original aadhar card. It both can be analyze it. Only authorized user can be authorized finger print and iris scan otherwise duplicate user displayed. In the C4.5 algorithm using accurate prediction on the original finger prints and iris scan, to store the aadhar database so to avoid fake aadhar card.

F. Back Propagation
In normal approach used for processing is called propagation. The back propagation algorithm requires long training times and extension testing and retraining of parameters, such as the number of hidden neurons learning rate and momentum, to determine the best performance.

Back propagation algorithm describes the first author name entered then it perform checking phase. It contains two categories demographic data and biometric data .Incase any one data checking means perform in reverse retraining process.
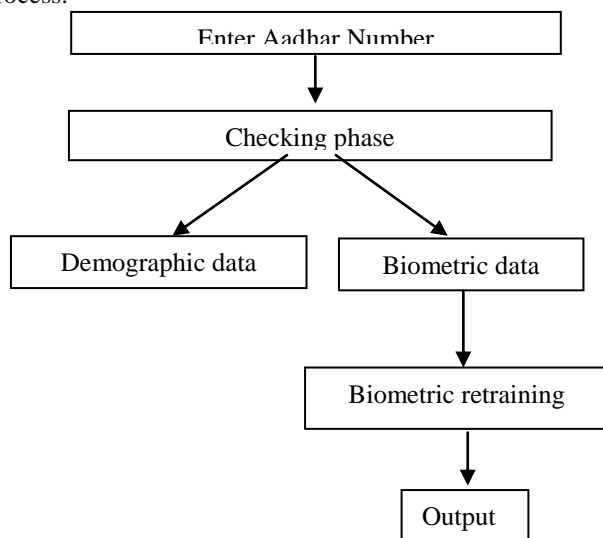

Figure :3 Checking phase by back propagation algorithm.

In this algorithm using aadhar card to biometric details refers to the iris, finger prints and facial photograph included. In the demographic details to be checked the order. Back propagation algorithm used in retraining process in demographic or biometrics any database checked it.

IV.COMPARATIVE STUDY

Comparative study of classification algorithms are naïve Bayesian, C4.5, back propagation used, and compare which one is best suitable for in this paper.

Table 1. COMPARATIVE STUDY

Algorithm	Problem Detection	Time Complexity
Naive Bayesian	Missing values	0.3
C4.5	Accurate prediction	82
Back propagation	Reverse retraining	0.42

The Navie bayes algorithm within overall time complexity is 0.75-0.3. C4.5 algorithm to problem detection the accurate prediction.

The time complexities reduce in the algorithm of naïve Bayesian and time complexity is 0.3. So in this algorithm is best suitable for fraud detection.

V.CONCLUSION

In this paper describes aadhar card fraud detection using some fraud detection method analysis will be carried out and to find some solutions and research will be directed towards creating a unified framework that trains with a specific and effective detection and even prevention is possible. Classification algorithm used in aadhar card some frauds are detected. Biometric details have some frauds to avoid the c4.5 algorithm finally some fake aadhar card reduced.

REFERENCES

[1] Mark A. Beyer and Douglas Laney. "The Importance of 'Big Data': A Definition". Gartner, 2012. For book

[2] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," interactions, vol. 19, no. 3, pp. 50–59, May 2012. For journal

[3]Pang -Ning Tan, Vipin Kumar,Micheal Steinbach, "Introduction to data mining", First Edition, 2012 For book

[4] D. Fisher, R. DeLine, M. Czerwinski, and S. Drucker, "Interactions with big data analytics," interactions, vol. 19, no. 3, pp. 50–59, May 2012 For journal

[5] Jiawei Han, Micheline Kambar, Jian Pei, "Data Mining Concepts and Techniques" Elsevier Second Edition. For book

[6] B. Thuraisingham, L. Khan, M. Awad, and L. Wang, Design and Implementation of Data Mining Tools. Florida, USA: Auerbach Publication, 2009 For book