Benami Identification Using Mining with Bigdata Analytics

Sonia Jenifer Rayen[1], Divya M[2], Farheen Tasneem Z[3], Nivetha M[4]
[1] Assistant profesor,
Department of information technology,
Jeppiaar Institute of technology,chennai.
soniya@jeppiaarinstitute.org
[2] Jeppiaar Institute of technology,chennai.
divyamohan2525@gmail.com
[3] Jeppiaar Institute of technology,chennai.
farheen_216@yahoo.in
[4] Jeppiaar Institute of technology,chennai.
nivetham01@gmail.com

Abstract: **The actual meaning of Benami is without a name. The word "Benami" means buying property in the name of another person. However, when you are making a purchase, you are not purchasing a plot on your name. You are purchasing this plot on the name of close relatives. The property purchased like this will be considered as a Benami property. The holder of the property is known as benamidar. In this paper, we introduce an Uniform Cost Search algorithm for efficient searching of benamidar. The benamidar is a person who has the movable and immovable, tangible, intangible, any right or interest, or legal documents asserts of another person in their name. As such, even gold or financial securities including bank accounts could qualify to be Benami. To identify the benamidar, the details of Aadhar card, the details of PAN card, the details of asserts and details of Bank are filled and stored. The interlink of Aadhar card, PAN card, asserts and Bank account is done. The mining of details is done by generating family tree and by pruning the person is being extracted. Each property is linked with Aadhar card so that on examining the details will help to find the benamidar by the Income Tax Officer.**

Keyword: Benami, Benamidar, Mining, Pruning

1.INTRODUCTION

The properties are generally purchased for the direct or indirect benefit of a person paying money for the purchase. The properties purchased on fictitious name. The owner of property is missing or not traceable. The property owner does not know about the transaction. The money is paid from black money and PAN card number is not provided. Incase if beneficiary and owner of a property are not same. The anti-corruption strategy aim at measures which make it difficult for the corrupt to spend or invest the ill gotten wealth. Most of the wealth in India are get invested in benami immoveable property, gold and jewelers, high value goods and other conspicuous consumption. The Benami Transactions (Prohibition) Act, 1988 (Benami Act) prohibits benami transactions and also it provides for government acquisition of the property held benami. Implementation of the Benami Transaction Act needs to be made more effective.

What are the exclusions for the Benami properties?

There is some exclusion for the Benami properties. This exclusion includes –If the property is purchased in the name of wife or unmarried daughter, it is not Benami property. If a property is held by wife or any other relatives with known fund it is excluded from

Benami properties. In case of Hindu Undivided Family (HUF), if head of Karta decide to buy a property for the benefits of all trustees. Religious or charitable properties are excluded.

How Benami Property will be identified?

In order to identify Benami properties, authorities consider following factors. Authorities will look at the source of fund for purchasing the property. Tax on the fund is paid or not. If the property is purchased on someone's name what is the purpose behind buying the property.The authority will find out who is holding a document of the property.If the property is bought the authority will check whether the tax is paid for the property or not. The authority will also check that actual possession of the property is with whom.

2.LITERATURE SURVEY

2.1 A Weighting-Based Local Search Heuristic Algorithm for the Set Covering Problem

In [1] A Literature Review on A Weight-Based Local Heuristic Algorithm for the Set Covering Problem Chao Gao, Thomas Weise, Jindong L, proposed a framework to perturb the candidate solution under which the best objective value found

during the search, a weighting scheme and several search strategies are adopted to help escape from local optima and make the search more divergent. The effectiveness of algorithm is evaluated on a set of instances from the OR-Library and Steiner triple systems.A very effective heuristic for SCPs are based on the linear programming relaxation (LP) or Lagrangian relaxation. In [1] introduced a local search heuristic algorithm based on weighting. The effectiveness of our algorithm has been tested on the instances from OR-Library and the STS. The experimental results show that new algorithm is comparable with 3-FNLS on instances from OR-Library, so it can find all the optima or BKS in very short times.

2.2 A comparison between Bee Swarm Optimization and Greedy Algorithm for the Knapsack Problem with bee reallocation

In [2] this paper a rough comparison for a search is done. The comparison with Bee Swarm Optimization and Greedy Algorithm for Knapsack Problem is done. Metaheuristic algorithms are now considered among the best tools to find good solutions with a reasonable investment of resources. This metaheuristic has proved its ability to deal with very complicated optimization and search problems. The collective and social behavior of living creatures motivated researchers to undertake the study of what today is known as **Swarm Intelligence**

The Bee Algorithm or BA, is also part of the Swarm Intelligence and this mimics the honey bees and their foraging behavior. This algorithm is based on a random search on the neighborhood for combinatorial and functional optimization.

3.PROPOSED SYSTEM

A. The unique identification project presents an opportunity to find the benamidar effectively. Providing the unique identity number and necessary document should be made compulsory for buying the property. The necessary document includes Aadhar card, pan card, land survey number and bank account details should be provided while purchasing any kind of properties. The detail of the person is being collected and stored in the database. On buying of property the details of the person who buys the property is being checked with the details of the documents whether the person buys the

properties in his own name or someone else name. The property details should be linked in the Aadhar card which is made compulsory by the government makes the Income Tax Authorities to find the Benamidar easily. This helps in avoiding the black money and make people to show off the immovable properties in account. In our proposed system, we store the details of all persons in database where database is not sufficient for storing the large number of datasets so we use big data for storing the huge datasets. By storing these data's, we are able to generate a family tree of a person with their details in order to find the Benamidar. With the help of family tree and the details of each person in the family tree, the Authorities will find whether the purchaser can able to buy the property according to the payable Income Tax norms. For mapping the details of the persons in the family tree we use Iteratively Reweighted Least Squares (IRLS)algorithm to perform tree pruning process. This algorithm is efficient for finding the Benamidar.

4.ALGORITHM

4.1 Iteratively Reweighted Least Squares (IRLS)

The method of **iteratively reweighted least squares** (**IRLS**) is being used to solve optimization problems. IRLS is used to find the maximum likelihood estimates of a generalized linear model. It is also used to find in robust regression to check an M-estimator.M-estimator is the way of mitigating the influence of outliers in other words normally-distributed data set. Consider an example by minimizing the least absolute error rather than the least square error.

The scope of application is said to be iteratively reweighted least squares to statistical estimation problems and is considerably wider than is generally appreciated. It results in extending beyond the exponential-family-type generalized linear models to other distributions, also to non-linear parameterizations, and to dependent observations. Various criteria are being estimated other than maximum likelihood, including resistant alternatives, may be used. The algorithms mostly are numerically stable, easily programmed without the aid of packages, and highly suited to interactive computation.

The IRWLS fitting Algorithm
1. Start with w_i,for example $w_i = 1$
2. Use least squares to estimates the value for

3. Use the residuals to estimates the value for °, by regressing x on ^² 2.

4. Recomputed and go to 2

Uniform Cost Search (UCS)

Breadth-first is only optimal if step costs is increasing with depth (e.g. constant). Can we guarantee optimality for any step cost?

Uniform-cost Search: Expand node with smallest path cost g(n).

It modifies BFS (Breadth First Search) by always expanding the lowest cost node on the fringe using path cost function g(n) (i.e. the cost of the path from the initial state to the node n).

Nodes maintained on queue in order of increasing path cost.

The algorithm is identical to the general graph search algorithm except for the use of a priority queue and the addition of an extra check in case a shorter path to a frontier state is discovered.

The data structure for frontier needs to support efficient membership testing, so it should combine the capabilities of a priority queue and a hash table

Function Uniform-Cost-Search(Problem) returns a solution, or failure

node<-a node with STATE=problem. Initial-State, Path-Cost=0

Frontier <- a priority queue ordered by Path-Cost, with node as the only element

Explored<- an empty set

Loop do

If EMPTY(frontier) **then return** failure

Node <- (pop)frontier

If problem. GOAL-TEST (node. STATE) then return SOLUTION(node)

add node. STATE to explored

for each action in problem.

ACTIONS (node. STATE)

 do

Child=CHILD.NODE(problem,node,action)

If Child. STATE is not in explored or frontier,

 then

Frontier<-INSERT(child, frontier)

Else if

child.STATEis in frontier with higher PATH-COST then

Replace that frontier node with child

IMPLEMENTATION:

fringe = queue ordered by path cost Equivalent to breadth-first if all step costs all equal.

Complete? Yes, if step cost $\geq$ Ɖ

(otherwise it can get stuck in infinite loops)

Time? # of nodes with path cost $\leq$ cost of optimal solution.

Space? # of nodes with path cost $\leq$ cost of optimal solution

Example hand-simulated search: Search tree method
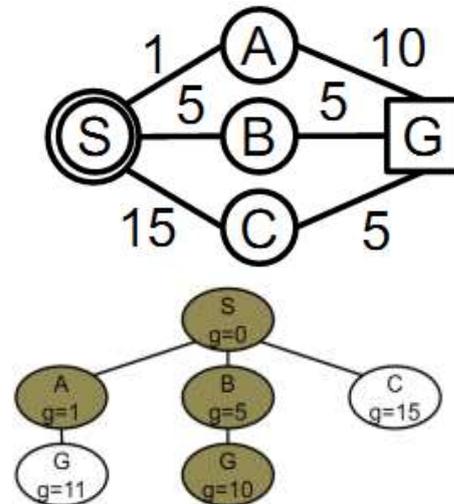
Order of node expansion: S A B G
Path found: S B G
Cost of path found: 10

6. BLOCK DIAGRAM
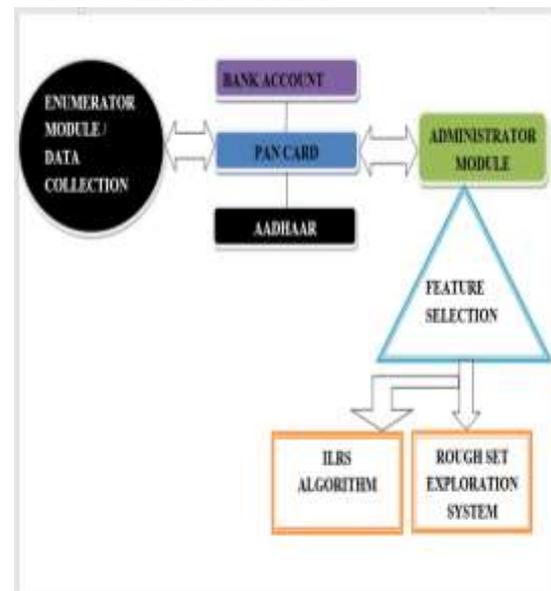
In the above diagram, the first module contains the enumerated data were the details of the persons are being stored in the system. The enumerator is a person who fills the data with the details of bank account, Aadhar card details and Pan card details.

The enumerator provides unique username and password for the person where the details of persons are being maintained in a database. The Enumerator will be given with national ID. The bank account details, Pan card details and Aadhar card details are being interlinked. The administrator in the diagram manages the entire

system. The administrator provides privileges to the enumerator and allows the access of system to the authorities who have rights. The administrator creates census rights so that the security of accessing the system is protected and details are also protected.

The Benami of a person or Benamidar is being found out using Mining technique. The feature selection module makes to eliminate the irrelevant data as we generate a family tree to find the Benamidar. The more relevant details for the person is being pruned so that it will help the authorities in finding the persons buying properties by not paying income tax and does not comes into their account. This selection helps in creating an efficient learning module where the data are found which are relevant to the search.

Next deals with the classification process where each object are being divided under some criteria so that the data in the system are being made relevant and the search is being made more efficient by the Tax Officers.

Since we already knew the concept, classification is known as supervised learning so that the process goes on observation and does in a systematic format. We use ILRS algorithm, this algorithm helps in reducing the data sets which means that large number of datasets are being reduced by relevant datasets which saves much of time in retrieving the relevant data.

The RSES is an exploration software and used as a knowledge obtaining database. This derives the decision rules from the available information system that relate the classes of condition attributes with classes of condition attributes.

7. SYSTEM ARCHITECTURE

8.CONCLUSION

The result of our proposed system is to find the Benamidar using the mining and pruning technique. The algorithm used by our system is ILRS which is an efficient technique to find out the irrelevant data for the given criteria. This reduces the time and complexity of finding the relevant datasets in the large number of datasets are made easier. At earlier stage Benami property identification are not easier to find which takes years in identifying the Benamidar. This made in the allocation of black money in one particular corner where it stops money flow over the world.

As said before the movable or immovable asserts are bought by the other people name whom definitely be a person under their family tree (the person might be is nearby friends, far away friends, nearby relatives, far away relatives). The family tree is created with the entire detail of the person his/her history, their identification number etc.

To overcome this Benamidar or to overcome the accumulation of the black money at a single spot our Honorable Prime Minister has taken a great step of Demonetization and the next huger step is to find Benami. Though there were lots of acts passed and lots of action taken against Benami Identification, our proposed system makes more efficient for the Authorities to identify the Benamidar in the system efficiently. Thus, the result of our proposed system "Benami Identification" is finding Benamidar.

REFERENCE

[1] Chao Gao1, Thomas Weise2, Jinlong L3 "A Weighting-Based Local Search Heuristic Algorithm for the Set Covering Problem"

[2] Marco Aurelio Sotelo-Figueroa, Mar´ıa del Rosario Baltazar-Flores, Juan Mart´ınCarpio, Victor Zamudio "A comparation between Bee Swarm Optimization and Greedy Algorithm for the Knapsack Problem with bee reallocation".

[3] C. Ravazzi; E. Magli 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[4] M Prakash, S Karthika, S Ramanani, "An Analysis for Classification of Soil Data in Data Mining using R", International Journal of Control Theory and Applications, 10 (2), pp 91-98, 2017.

[5] Mohan, Prakash, and Ravichandran Thangavel. "Resource Selection in Grid Environment Based on Trust Evaluation using Feedback and Performance." *American Journal of Applied Sciences* 10.8 (2013): 924.

[6] Prakash, M., and T. Ravichandran. "An Efficient Resource Selection and Binding Model for Job Scheduling in Grid." *European Journal of Scientific Research* 81.4 (2012): 450-458.

[7] Prakash, M., R. Farah Sayeed, S. Princey, and S. Priyanka. "Deployment of MultiCloud Environment with Avoidance of

DDOS Attack and Secured Data Privacy." *International Journal of Applied Engineering Research* 10, no. 9 (2015): 8121-8124.

[8] Annamalai, R., J. Srikanth, and M. Prakash. "Integrity and Privacy Sustenance of Shared Large Scale Images in the Cloud by Ring Signature." *International Journal of Computer Applications* 114.12 (2015).

[9] Mohan Prakash, Chelliah Saravanakumar. "An Authentication Technique for Accessing De-Duplicated Data from Private Cloud using One Time Password", International Journal of Information Security and Privacy, 11(2), 1-10, 2017.

[10] An improved IRLS algorithm with double-threshold sigmoid penalty Renhuan Ding; Jun Shi; Gao Xiang; Xiaoling Zhang 2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)

[11] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," Multiscale Model. Simul., vol. 4, pp. 490–530, 2005.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, pp. 2080–2095, 2007

[13] Jungang Yang ; Thompson, J. ; Xiaotao Huang ; Tian Jin ; Zhimin Zhou; Segmented Reconstruction for Compressed Sensing SAR Imaging, IEEE Transactions on Geoscience and Remote Sensing, Volume: 51 , Issue: 7 , Part: 2, 2013 , pp. 4214 – 4225.

[14] Gorodnitsky and B. Rao, Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm, IEEE Trans. Signal Process., vol. 45, no. 3, 1997, pp. 600-616.