

# Prediction and Retrieval of Information in Big Data Technology with Data Warehouse

C. B. David Joel Kishore\*<sup>1</sup>, Dr. T. Bhaskara Reddy<sup>2</sup>

\*<sup>1</sup>Research Scholar, Department of Computer Science, Rayalaseema University, Kurnool, Andhra Pradesh, India

<sup>2</sup>Professor, Department of Computer Science & Technology, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, India

## ABSTRACT

The big data technology with the data warehouse provide better public services analytics and it can help governments make quicker and more effective policy, investment and infrastructure decisions. The data warehousing and the big data is a storehouse of land record information. The need for storing and retrieving land record data is increasing progressively due to population growth, education, political analytics etc. In order to manage these kinds of information, Government need to use efficient and large database for storage purposes and retrieve correct information's. Besides, the problems may occur due to some challenges while retrieving a person's details from the large data sets. Also maintenance of enormous quantity of data is not a simple task. In order to eliminate such limitations, we proposes a novel Fuzzy-Cat Swarm Optimization (FCSO) method to analyse, store and retrieve the land record information accurately. This proposed method consist of, Artificial Neural Network (ANN) that classifies the input land record data for organizing the information as different classes in the database. Then, the proposed method uses a mongo database to store large amount of land record information. Moreover, it provide easy maintain ace of data, updating of land records and security of the system. Besides, the FCSO approach is used to attain the accurate retrieval results for user's queries. Thus, the users can access their records easily by means of fuzzy optimized approach. Therefore, the land record data maintenance and decision making for retrieving information can be easily and efficiently processed using this proposed system. Finally, the performance of the proposed method is evaluated for the land record information retrieval results.

**Keywords:** Big data, data warehousing; land record data, artificial neural network, mongo database, fuzzy, cat swarm optimization.

## I. INTRODUCTION

The system of land records management varies across states, depending on factors such as historical evolution and local traditions. For the registration and maintenance of such land records, it is difficult to ensure that survey maps, textual data, and registration records match with each other and are updated [1]. Though, the big data technology with data warehouse system provide an efficient approach for land record classification and retrieval of classified land record

data's. Data analysis and storing the amount of information are utilized in the structure of the data warehouse [2]. All data relevant to the management of the organization are placed in data warehouse and it acts as a central repository which the prominent information and knowledge need to effectively manage the organization [3]. For integrating the data and development, management, operational methods, and practices the data warehouse is important that define how these data are collected, integrated, interpreted, managed, and used. Data warehouses

include experience-based organizational knowledge and it should be understandable one [4]. For the better decision making data warehouse is processed to keep the large database. The big data technologies used for storing the infinite number of data and also we make a good decision [5]. To analyse the large volume of data and extract the necessary information or knowledge for future action using big data.

In this framework the detailed data warehouse to provide the right information in a timely manner to policy-providers, decision-making users and peoples to support them in performing their jobs more effectively and efficiently. To enhance the development, they prepared and submitted a project to the Department of Electronics (DoE), New Delhi for a prototype system [6]. In their project proposal, they used the Government of Andhra Pradesh (GoAP) as a test bed for the proto type system mainly because they already had the necessary data for the system in electronic form. After that a common database warehouse is maintained to deal with the information related to the common people. There has been a lot of computerization projects proposed by various state Governments as well as national level planning organizations [7]. The resulting database was for every household in the state with the particulars of each member of the household. This database is known as MPHS (Multi-Purpose Household Survey) database and is available with the Government of Andhra Pradesh in electronic format. We decided to use it for person related information.

They predominantly used stand-alone applications catering to only a limited functions in this level. The information generated by these standalone database applications arrived in sporadic and heterogeneous manner [9]. This makes it difficult for the decision making users and policy providers to access current level data with a high granularity. As decision-makers cannot extract useful information directly from these systems due to non-availability of suitable tools. The

IT staff cannot produce this decision support information in timely manner due to non-uniformity of data and lack of suitable tools [10]. Therefore these works motivate us to improve the database maintenance work. For effective data processing and storage data warehouse techniques are handled. To achieve effective search and separate database maintenance big data is included in this work.

Land Records itself is a valuable asset and a generic expression that can include records such as, the register of lands, Records of Rights, tenancy and crop inspection register, mutation register, disputed cases register, etc [11]. It can also include geological information regarding the shape, size, soil-type of the land; and economic information related to irrigation and crops. So Big data concept is used for collecting more information [12]. Big data describes the large volume of data both structured and unstructured that inundates a business on a day-to-day basis. Land administration involves recording, processing and dissemination of information about the ownership, value, and use of land. It mainly depended on location and the growing population. The land records management varies according to states, depending on factors such as historical evolution and local traditions etc [13]. Such information can be broadly classified as details of the property, spatial records, and transaction records.

Nowadays, the land ownership can be determined through a set of documents. Poor land records also affect future property transactions. It became more difficult to access land records when data was spread across departments and has not been updated. One had to go back to several years of documents to find any ownership claims on a piece of property and this may cause delay [14]. Furthermore, it was difficult to update and ensure that the survey maps, textual data, and registration records match with each other because of multiple agencies responsible for registration and maintenance of records. In addition,

citizens had to approach several agencies to get complete information on land records [15]. Poor maintenance of land records also led to inaccuracy in them. The household survey includes the community survey, housing survey, population survey, expenditure survey etc. In order to retrieve information from such kind of data is a challenging task.

Although, there are different method for better decision making system by using warehousing and big data storage system. It maintain the MPHS data, land records data of the Andhra Pradesh Government. This system uses those information as inputs for data maintenance and decision making. By taking the entire data from the government it is hard to maintain the separate data section for each and every individual person. So this work concentrate on multiplies the data and also stored in a single database [16]. Input data is classified as per individual details and stored into the database. Different sources of collected land records and person details are converted into uniform format. So that the data are easy to separate under one particular category. Those conversion of data and classification can be done by using some classification techniques. They are:

Naive Bayes Classifier mainly used for simple and complete system for sentiment mining on large datasets the average accuracy stayed below 82% .[17]. LIME classifiers, have made significant improvements to the performance of base classifiers and standard ensemble Meta classifiers [16]. Neuro-Fuzzy Classifier, it not only help to reduce the dimensionality of large data sets but also to speed up the computation time of a learning algorithm and therefore simplify the classification task [18]. For land record storage system, some of the techniques are used as: data mining base algorithm for the storage location of data in the warehousing [19]. The cloud based architecture for storing large amount of data in warehousing [20]. It was also used for

mapping the land areas, agriculture field by crop management system which were to be deployed and validated [21]. Then for the land record data retrieval varying methods are used such as query based approach for information retrieval system in the cloud storage of warehousing systems [22]. The vector space model was mainly used in data warehousing and cleaned for the purpose to provide accurate and valid data [23]. Though, the above techniques are effective for land record classification and retrieval, but it not provide better and accurate result to individuals. In order to overcome from such limitations, the proposed approach is introduced.

This manuscript proposes a new efficient method for land record information retrieval by in big data technology with data warehouse system. A new method for better decision making system has been proposed by using warehousing and big data storage system. Using the proposed system we maintain the MPHS data, land records data of the Andhra Pradesh Government. By taking the entire data from the AP government, it is hard to maintain the separate data section for each and every individual person. So, this work concentrates on classifying the land record data and then stored in a single database and retrieve the corresponding information. Classifying and retrieving the land record information accurately from a large source of data is a challenging task. So, we propose a novel efficient FUZZY-CSO approach to classify and retrieve the land record information accurately. This proposed framework allow the decision makers and policy planners to access their significant data. As a result, the proposed system improves the accuracy and provides high integrity assurance. The major contribution of the proposed work are:

- 1) We modelled an FUZZY-CSO approach to predict and retrieve the information accurately.
- 2) We intended to use an ANN classification model to arrange the database for easy maintenance and access for individuals.

- 3) We develop a novel fuzzy based optimization approach for enabling easy maintenance and prompt updating of land records. The Cat Swarm Optimization (CSO) algorithm will allow the users to easily access their records optimally.
- 4) The experimental results outperforms the proposed framework that allows the decision makers and policy planners to access their relevant data it also improves the accuracy of the proposed system and provides greater integrity assurance.

The remaining of the paper is organized by the following sections. Related work is described in section 2. Section 3 defines the proposed methodology in detail. The Experimental result is discussed and comparison of various methods are explained in section 4. Finally, Section 5 provides the conclusion.

## II. PROPOSED PREDICTION AND RETRIEVAL OF INFORMATION

The three main important sections of proposed approaches are classification of land record information, database storage and the efficient land record information retrieval. This proposed framework make use of land record data and the MPHS data of an Andhra Pradesh government. Both the records are collected from different source of same government dataset and those data are converted into the uniform format. It classified according to ROR, spatial records, transactions records of each land. After that, the collected information are classified by means of ANN classifier. It is an efficient classifier that provide accurate results after classifications. After the effective classification, the information are stored in the mongo database. In addition to, the mongo database is an efficient database that can store large amount data other than any database. Moreover, it's a document-oriented databases, so the land record information of each individual is stored in an

effective manner. Subsequently, the classified and stored information are retrieved based on the FCSO approach. The fuzzy optimized ranking function effectively retrieve the user's information. Besides, the FCSO approach is used to attain the accurate retrieval results for user's queries. Thus, the users can access their records easily by means of fuzzy optimized approach. Therefore, the land record data maintenance and decision making for retrieving information can be easily and efficiently processed using this proposed system. The below Figure 1 represents the process flow of the proposed system.

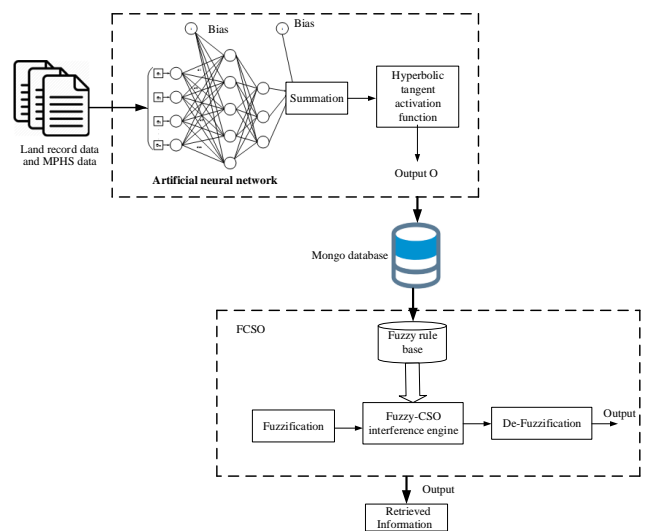


Figure 1. Architecture of Proposed System

### A. Classification of Land Record Data

In this proposed framework, the ANN is used for classifying the land record informations. In the classification, the land record data and the MPHS data are consider from different sources of same dataset. During processing the information were not in a proper form. Consequently, the users can't easily access their informations. Besides, it is mainly used in the classification problem. The ANN is a non-parametric method that does not take advantage of the statistical frequency distribution of the data. Among numerous solutions of neural networks, the most widely used algorithm in classification problem, which consists of one input layer, at least one hidden layer and one output layer. Each layer consists of

several neuron nodes and all the nodes in a layer are connected to the nodes in the adjacent layers, but there are no interconnections between nodes with the same layer. Each interconnection between nodes carries an associated weight, and each node computes the weighted sum of the inputs and passes the sum through an activation function that provides the output value of this node. The important objective of ANN is to stimulate a system to act various computational tasks such as pattern recognition and classification, optimization and data clustering.

At first, the collection of input land record data as  $n_1, n_2, \dots, n_m$  is multiplied by connection weights  $w_1, w_2, \dots, w_m$ . Then, the products are summed and fed to the activation function.

The output of Summation is given,

$$K_{in} = \sum_i^n n_i \cdot w_i \tag{1}$$

The output of the activation function as,

$$Z = F(K_{in}) \tag{2}$$

Summation process  $S = \sum_i^n n_i \cdot w_i$  bias can be considered as,

$$S = \sum_{i=1}^n n_i \cdot w_i + BIAS \cdot w_i, \quad \text{when} \\ BIAS=1 \tag{3}$$

The above equation represents the sigmoid activation function and hyperbolic tangent activation function, which is a sigmoidal activation in the form of a hyperbolic tangent. The hyperbolic tangent is a ratio of corresponding hyperbolic sine and hyperbolic cosine function via in Figure 2.

The nodes in the network represent the number of inputs from land records. The output of the input layer is given as the input to the hidden layer and the hidden layer consist of two or three layers. The main function is to encode the data and maps it to the output. Output layer delivers the classified result and network has been trained to provide efficient outcome. The correct class for each node is known and the output mode assign the value as 0 and 1 based on the classes. The iteration is repeated until get the correct

value as the output. The back propagation layer is also known as feed forward layer which has input, output and hidden layer.

Hyperbolic tangent activation function,

$$o = \tanh(s) = \frac{\tanh(s)+1}{2} \tag{4}$$

Though, for the weight changes of hidden layers and the output layers the back propagation algorithm is used. In order to reach the input layer, the local gradients of each land record data is going backward direction based on the back propagation algorithm. Moreover, the delta-rule is used by back propagation algorithm. During classification, the error might be occurred in between the network output and the desired output. Besides, it need to match the output of the desired information. Further, by the back propagation rule, adjust the weight by adding new weight for the output layer. After that, by using the momentum coefficient calculate the current weights and the pervious weight of the system.

After calculating the coefficients, the mapping function is used, in which it map the input variables and the output variables based on the new data that predicts the output variables. As a result, the algorithm achieves an acceptable performance. Therefore, for the better performance, the process must be done more and more. Finally, the classification output is match each other and then updated to the storage system. Besides, the classified information is stored in the mongo database.

### B. Data Storage

In our proposed work, the mongo database is used as the data storage system. It store large amount of classified land record data's. Besides, it store multiple amount of data for easy maintainece and promoting land records and the security system. It provides more simple and reliable approach to meeting diverse application needs for data. With Mongo database flexible storage architecture, the database automatically manages the movement of data between storage engine technologies using native replication.

By an expressive query language, documents can be queried in many ways – from simple lookups to creating sophisticated processing pipelines for data analytics and transformations, through to faceted search etc. With a flexible storage architecture, application owners can deploy storage engines optimized for different workload and operational requirements. Thus, this system has many advantages for storing multiple files in different fields. Thus, the user can access their corresponding information from desired field of the mongo database.

#### **i) Database model**

The mongo database model consist of group of land records. The model of the mongo database is a record database model. With mongo database, the organizations can address diverse application needs, hardware resources, and deployment designs with a single database technology. Mongo DB can be extended with new capabilities, and configured for optimal use of specific hardware architectures. This approach significantly reduces developer and operational complexity compared to running multiple databases to power applications with unique requirements. Users can control the same mongo DB query language, data model, scaling, security and operational tooling across different applications, each powered by different pluggable mongo DB storage engines. Besides, it provide better scalability for storing all the classified spatial, textual, and the transaction details of the land record information in a single database.

#### **ii) Data as Document**

The mongo DB records number of the document in a single particular database. Mongo DB documents are closely aligned to the structure of objects in the programming language. This makes it simpler and faster for developers to model how data in the application will map to data stored in the database. Mongo DB documents tend to have all data for a given record in a single document, whereas in a

relational database information for a given record is usually spread across many tables. The data structure combine multiple rows and the column such in a table that represent as user, commands, articles, sections and tags. In Mongo DB the data could be modelled as two collections, one for users, and the other for articles. In each blog document there might be multiple comments, multiple tags, and multiple categories, each expressed as an embedded array. As a result of the database document model, data in Mongo DB is more localized, and reduces the need to JOIN separate tables. Therefore, the database possess higher performance and scalability across commodity hardware as a single read to the database can retrieve the entire document.

#### **iii) Mapping Coordinator**

The mapping coordinator is an important asset in the mongo database. It mainly used for implementing and maintain the hardware and the software system to declare reliable, compatible and user-friendly operations. Besides, the mapping coordinator able to retrieve the information accurately in a robust way. Consequently, if the user needs any authenticated data about the organization, social or use, then the mapping coordinator check whether the query is authorized user or not. After all the confirmation procedures, the system provide the information to the intended user. Thus, it provide security criteria to the users.

### **C. Fuzzy cat swarm optimization for Information Retrieval**

In our proposed framework, the fuzzy cat swarm optimization technique is applied. It mainly used for retrieving the land record information using a specific set of fuzzy rules. Besides, it operates as an if-then conditions. The membership function is mainly based on the relevant document and its value between 0 and 1. Ranking functions recovered a ranking value for all rows in a separation. A Ranking is a process of ordering the results of the query

according to the measures relevance to the corresponding user.

The ranking process is obtained using a weighting function

$$(tf \times idf),$$

Which conveys how relevant document 'c' is for query 'p'.

Let us assume the set of documents 'C' in mongo database with relevance numerical weights values

$$b_i = \{b_1, b_2, \dots, b_n\} \quad (5)$$

Where  $b_i \in \{0,1\}$  as normalization interval prompts a relevant document  $c_i$ .

Here '1' is the maximum relevance weight value corresponding to 'high relevant' and '0' represents 'irrelevant value'. The matrix representation for the Mongo document is represented as,

$$C = \begin{bmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & c_{22} & c_{23} & \dots & c_{2n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ c_{M1} & c_{M2} & c_{M3} & \dots & c_{Mn} \end{bmatrix} \quad (6)$$

The rank list for the retrieved documents are given according to the weight value  $b_i$  is calculated as

$$w = (tf \times idf) \quad (7)$$

Where,  $tf$  is the term frequency in the query document pair and the fraction of retrieved documents  $idf$  is represented as

$$idf = \log(ni/M) \quad (8)$$

Where,  $ni$  is the number of documents retrieved and  $M$  is the total number of documents.

The Ranking function  $E$  consists of an ordered set of ranks. Each rank consists of relevance weight value  $b \in \{0,1\}$  where  $b$  represents the relevance numerical weights of the retrieved documents. Each retrieved document is assigned with an ascending rank number.

$$E = \{(1, b_1), (2, b_2), (3, b_3), \dots, (n, b_n)\}$$

Where,  $b_1 > b_2 \dots b_n$

The FCSO approach mainly used to retrieve the informations accurately by means of specific set of rules. The user can access and retrieve the land record data accurately and in a robust way. On the other hand, the CSO method reduce the optimization

problem while retrieving the informations. As, it was mainly composed in two models such as tracking mode and seeking mode which represents two different procedures in the algorithm. The CSO algorithm has the finer attainment function to minimize the problems that are based on other optimization algorithms. In CSO, cats are used to indicate the solution. The optimization problem is solved by monitoring the behavior of cats like resting, moving, hunting etc. The two modes that are used to monitor the behavior is seeking mode and tracing mode. Both the modes are combined together based on the definition of user to obtain effective performance.

**Seeking mode:** The seeking mode determines the behavior of cats mainly while resting or looking around by sitting in its own position. The cats make a slow movement if it recognize a prey or any danger situation. The seeking mode operates under four factors for determining the memory size, dimension for mutation, for denoting the varied dimension and to decide the point of the cat to move to next position.

$$P_i = \frac{|F_{si} - F_{sb}|}{F_{smax} - F_{smin}}, 0 < i < j \quad (9)$$

If the fitness function is to find the minimum solution then consider

If the fitness function is to find the minimum solution then consider  $F_{sb} = F_{smax}$  otherwise  $F_{sb} = F_{smin}$ .

**Tracing mode:** The tracing mode indicates the hunting behavior of the cats. The movement of the cats in this mode is faster than the seeking mode. The number of cats in the tracing mode is calculated by the mixture ratio (MR). After sorting the cats that are available in both modes, the cats will be positioned in new position.

$$V_{k,d} = V_{k,d} + r_1 \times c_1 \times (X_{best,d} - X_{k,d}) \quad (10)$$

Where  $d=1, 2, \dots, M$ ,  $X_{best,d}$  denotes the cat's position which holds best fitness function,  $X_{k,d}$  denotes the cat's position,  $c_1$  is constant,  $r_1$  denotes the random value.

$$X_{k,d} = X_{k,d} + V_{k,d} \quad (11)$$



The cat swarm optimization works in the following form,

**Step 1:** Create the N-number of cats that are applied in the search space from the result of the fuzzy-based ranking system.

Query vector  $P$  as  $P = [p_1, p_2, \dots, p_n]$

**Step 2:** The selected cats are placed in its own dimension with its gives the values in the order of highest velocity.

For N-cats( $N \times M$ ), M is the number of variables optimized in the range (0, 1).

**Step 3:** The fitness function represent the distance between the cats. Evaluate the fitness value of each N cats and the best fitness value of cat act as the global best (gbest).

Evaluate the documents according to their fitness function  $f(x)$ , keep the best fitness value.

$$f(x) = \max \left\{ \frac{1}{i} \sqrt{\sum_{j=1}^i w_j^2} \right\} \quad j = 1, 2, \dots, n$$

**Step 4:** If the cat is resting then it is in seeking mode.

**Step 4.1:** Seeking memory pool-It determines each cat's seeking point and according to the rules the cat may choose the next point.

**Step 4.2:** Seeking range of the selected dimension- The mutation process is done for the selected dimension.

**Step 4.3:** Counts of dimension to change- It reveals all the varied dimension. The CDC has the out of M-dimensions of a cat if CDC=1, then the dimensions of CDC is randomly changed. Repeat the procedure for all cats, and evaluate the fitness value of each position. The best-fitted cat is retrained and the remaining will be discarded. Repeat it for all seeking mode cats.

**Step 4.4:** Self position considering- It is a variable which decides the cat to move to the next position.

**Step 5:** If the cat is hunting a prey, then it is in tracing mode and estimate fitness function.

$$V_{k,d} = V_{k,d} + r_1 \times c_1 \times (x_{gbest,d} - x_{k,d})$$

Where,  $x_{best,d}$  - Position of the cat, who has the best fitness value, and  $V_{k,d}$  is the velocity of  $cat_k$  on the  $d^{th}$  dimension,  $x_{k,d}$  is the Position of the cat, which random variable  $r_1$  in the range of (0, 1) and constant  $c_1$ .

$$x_{k,d} = x_{k,d} + V_{k,d}$$

**Step 6:** Re-pick the number of cats and set them into the tracing mode according to MR, then set the other cats into the seeking mode.

**Step 7:** Check the termination condition, if satisfied, terminate the program, otherwise repeat the steps 3 to step 7.

### III. EXPERIMENTAL RESULT AND ANALYSIS

In this paper, FCSO method is proposed to classify, storing and retrieving the land record and the MPHS data from AP dataset from different sources. The proposed technique provides the better performance to classify the land records of Andhra Pradesh government. The different district of the Andhra Pradesh government, land holders, and their properties information are included in the content input table. Each of the districts can be collapsed itself. So using ANN classifier for classify the data and each of the districts will be separated. Each of the individual information is separated using different kinds of function in ANN classifier for user convenient. It separates the different district name and the land holders name and their properties. Then



the Classification results are stored in the efficient mongo database. After that, the user required information is retrieved from the mongo database using the Fuzzy ranking CSO function. The experimental results and analysis are performed in MATLAB 2017b tool. It evaluates the following terms such as precision and recall, F-measure, efficiency, accuracy, and scalability.

The AP land record data set consists of 96, 58,090 documents of size 48 GB is stored in mongo database such as 87,08,469 ROR, 1,09,300 mutation register, 3,15,798 village map, 1,31,641 city survey maps, 1,24,590 encumbrance certificate, 1,51,926 property tax receipts and 1,16,366 sale deed documents with 17,563 attributes (number of villages) and six classes. The proposed approach is applied on each field of the database for user convenient retrieval.

The AP government includes 4,63,21,923 documents MPHS data, 28,196 attributes and 25 classes with database size of 65GB, which include reports, codebook, building material, expenditure, physical assets and loans, unemployment in men or women, growth rate of population region wise, land holding of people below certain level, crop wise area and cultivation trend, school age children etc.

**A. Precision and Recall**

Precision and the recall is an important evaluation metrics during information retrieval. The precision is defined as, fraction of retrieved relevant information to total retrieved information. As well as, the recall is a fraction of retrieved relevant information to total relevant data. The retrieved documents not retrieved documents are represented by true positive, false positive, false negative and true negative value.

**B. Accuracy**

The accuracy is the summation of True Negative Rate (TNR) and the recall. The TNR is also known to be specificity. The system achieves high accuracies when both precision and trueness are high. Accuracy

is the proportion of true results such as relevant retrieved and irrelevant not retrieved.

The existing systems are required for the following functions such as Precision, Recall, F-measure, and Accuracy.

**Table 1.** Performance Evaluation Of The Proposed System

Evaluation metrics	Land Record Data	MPHS Data
Precision (%)	87	85
Recall (%)	89	87
Accuracy (%)	99.28	99.85

Table 2 represents the performance evaluation of the land record data and MPHS data in terms of precision, recall, and accuracy for information retrieval. The land record information retrieval process consist of recall, precision and the accuracy of the test document required for the need of information. Therefore, the proposed method achieves high accuracy, precision and the recall rate. Thus, the user can retrieve the information accurately in a robust way.

**IV. CONCLUSION**

Nowadays, the development of infrastructure is growing as shifting of land use from agriculture to commercial, industrial, and residential. On the other hand, the land is an important asset in the world and its value depends on its location, and with growing population, and its demand keeps increasing, while its supply is limited. Furthermore, it was difficult to update and ensure that the survey maps, textual data, and registration records match with each other because of multiple agencies responsible for registration and maintenance of records. In addition, citizens had to approach several agencies to get complete information on land records. Poor maintenance of land records also led to inaccuracy in the overall system. The household survey includes the community survey, housing survey, population survey, expenditure survey etc. In order to retrieve

information from such kind of data is a challenging task. Therefore, the main purpose of this paper is to provide easy access, storage and retrieval of land-related information to the common user. In order to overcome such drawbacks, we proposed a novel FUZZY-CSO approach to predict and retrieve the land record information accurately. Then, the proposed method uses a mongo database to store large amount of land record information. Besides, the FCSO approach is used to attain the accurate retrieval results for user's queries. Thus, the users can access their records easily by means of fuzzy optimized approach. Therefore, the land record data maintenance and decision making for retrieving information can be easily and efficiently processed using this proposed system. Finally, the performance of the proposed method is evaluated for the land record information retrieval results.

## V. REFERENCES

- [1] Chawla R, Bhatnagar S. "Online delivery of land titles to rural farmers in Karnataka", India. In *Scaling up Poverty Reduction: A Global Learning Process and Conference 2004* May 25 (pp. 25-27).
- [2] Subramanian GH, Wang K. "Systems Dynamics-Based Modeling of Data Warehouse Quality". *Journal of Computer Information Systems*. 2017 Oct 20:1-8.
- [3] Cardon D. "Database vs data warehouse: A comparative review". *Health catalyst*. 2014.
- [4] Al-Debei MM. "Data warehouse as a backbone for business intelligence: Issues and challenges". *European Journal of Economics, Finance and Administrative Sciences*. 2011; 33(1):153-66.
- [5] Cheng Y, Kuang Y, Shi X, Dong C. "Sustainable Investment in a Supply Chain in the Big Data Era: An Information Updating Approach". *Sustainability*. 2018 Feb 4; 10(2):403.
- [6] Sinha PK. "Multipurpose, multilingual, multimedia, scaleable information warehouse for state level governance". *IETE Technical Review*. 2001 Nov 1; 18(6):443-7.
- [7] Latchem C. "Open and Distance Non-formal Education" in *Developing Countries*. Springer; 2017.
- [8] Ricardo Jorge Santos, Jorge Bernardino, Marco Vieira "24/7 Real-Time Data Warehousing: A Tool for Continuous Actionable Knowledge" *35th IEEE Annual Computer Software and Applications Conference*, 2011
- [9] Shi J, Bao Y, Leng F, Yu G. "Study on log-based change data capture and handling mechanism in real-time data warehouse". In *Computer Science and Software Engineering, 2008 International Conference on 2008* Dec 12 (Vol. 4, pp. 478-481). IEEE.
- [10] Ayhan, Samet, Johnathan Pesce, Paul Comitz, David Sweet, Steve Bliesner, and Gary Gerberick, "Predictive analytics with aviation big data", *Integrated Communications, Navigation and Surveillance Conference (ICNS)*, pp. 1-13, IEEE, 2013.
- [11] Khanna R, Khanna V. A Critical Appraisal of e-Governance in Rajasthan State with Special Reference to Tribal District Dungarpur, Rajasthan, India. *OmniScience: A Multi-disciplinary Journal*. 2018 Jan 15;7(3):23-6.
- [12] Torsten Priebe, Stefan Markus, *Business Information Modeling: A Methodology for Data-Intensive Projects, Data Science and Big Data Governance*, IEEE International Conference on Big Data (Big Data), 2015.
- [13] GPS traces, "IEEE Transactions on Intelligent Transportation Systems" 14, no. 1, pp. 113-123, 2013
- [14] Vijay Katkar, Surupendu Prakash Gangopadhyay, Sagar Rathod, and Aakash Shetty, "Sales forecasting using data warehouse and Naïve Bayesian classifier", *Pervasive Computing (ICPC)*, 2015 International Conference on, pp. 1-6. IEEE, 2015.

- [15] Walby K, Luscombe A. "Criteria for quality in qualitative research and use of freedom of information requests in the social sciences". *Qualitative Research*. 2017 Oct;17(5):537-53.
- [16] Schinasi LH, Auchincloss AH, Forrest CB, Roux AV. "Using electronic health record data for environmental and place based population health research: A systematic review". *Annals of Epidemiology*. 2018 Mar 21.
- [17] Bingwei Liu, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. "Scalable sentiment classification for big data analysis using Naive Bayes Classifier", *IEEE International Conference on*, pp. 99-104, 2013.
- [18] Jemal H. Abawajy, Andrei Kelarev, and Morshed Chowdhury, "Large iterative multitier ensemble classifiers for security of big data", *IEEE Transactions on Emerging Topics in Computing* 2, no. 3, pp. 352-363, 2014.
- [19] Ahmad Taher Azar, and Aboul Ella Hassanien, "Dimensionality reduction of medical big data using neural-fuzzy classifier", *Soft computing* 19, no. 4, pp. 1115-1127, 2015.
- [20] Diatchenko L. "Genotypic Data in Relational Databases: Efficient Storage and Rapid Retrieval". In *Advances in Databases and Information Systems: 21st European Conference, ADBIS 2017, Nicosia, Cyprus, September 24-27, 2017, Proceedings 2017 Sep 26* (Vol. 10509, p. 408). Springer.
- [21] Kriti Saroha, Anjana Gosain, "Multi-version Data Warehouse: A Survey", 2014.
- [22] Pavón-Pulido N, López-Riquelme JA, Torres R, Morais R, Pastor JA. New trends in precision agriculture: a novel cloud-based system for enabling data storage and agricultural task planning and automation. *Precision Agriculture*.1-31.
- [23] Li Z, Huang Q, Carbone GJ, Hu F. "A high performance query analytical framework for supporting data-intensive climate studies". *Computers, Environment and Urban Systems*. 2017 Mar 31; 62:210-21