Governing Scalable State Level Information Warehouse Using Big Data

[1]C. B. David Joel Kishore, [2]Dr. T. Bhaskara Reddy

[1]Research Scholar, Department of Computer Science, Rayalaseema University, Kurnool, AP, India.

[2]Professor, Department of Computer Science & Technology, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, India.

Abstract: The main repository of information is Data Warehouse (DW). The information can be stored at a particular level. The data warehouse collects recent and ancient data during the program for creating analytical reports for capability workers. The Mongo database for storing a wide amount of data is proposed in our methodology. In Mongo Database, Mapping Co-ordinator is included to supply user security. The input data and the information are stored in the Mongo database by the Support Vector Machine (SVM) classifier. The fuzzy based ranking function is used for efficient and easy usage of common people to retrieve the information from the database based on fuzzy rules. Data maintenance and decision making for retrieving information can be easily and efficiently processed using this proposed system.

Keywords: Data Warehouse, Support Vector Machine, Fuzzy Based Ranking Function, the Mongo database, Fuzzy rule.

I. INTRODUCTION

Data analysis and storing the amount of information are utilized in the structure of the data warehouse. 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 [1]. The information stored at the warehouse is in the scheme of derived views of data from sources [2]. 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, versatile [3]. To find situations requiring action and to know the situation and its causes they must supply information that enables managers [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]. So they proposed remarkable uses the bid data technology in the data warehouse. To analyse the large volume of data and extract the necessary information or knowledge for future action using big data [6]-[7].

In the population of 6 million that are utilized to store in big data technologies in existing the large-scale real-world taxi trace data of a big city. There the big data utilizing improved clustering algorithm (iterative DBSCAN) is presented to extract regions, respectively to the quality of the data. The best feature combination is achieved and the achievement of different features and classifiers is evaluated. The system considers only current addresses regions with pure land use and it does not consider regions with multiple land-use classes. So the Information of the passenger cannot be accessed clearly [11]-[12]. A fast Messy Genetic Algorithm used In Advance Hybrid Genetic Algorithm, a fast Messy Genetic Algorithm is used with the support vector machine (SVM) system was presented. The fmGA is used in the early prediction of conflict disposition in the initial stage of public-private partnership projects. The term of accuracy, precision, sensitivity, specificity, and the area under the curve are provided in the system. But the system has some classification problem so that doesn't give the efficient result [13]. Another huge data utilizing administration situated choice emotionally supportive networks and this framework given better outcomes, for example, the decrease in unit benefit costs because of increment in operational size (scale), diminishment in unit benefit costs because of increment in number of administrations being produced and given (extension) and lessening in unit costs because of increment in number of advantages are put through supply/request chain (speed). Be that as it may, the outcome isn't exact and does not effective [14]-[15].

The following sections organized the remaining paper. Related work is described in the section 2. Section 3 defines the proposed methodology in detail. In section 4, the experimental result is discussed. In the section 5 various methods comparison and analysis is explained. The Section 6 finally provide the conclusion.

II. RELATED WORK

Stefan Markus *et al.* [16] have proposed a consolidated technique to structure and depict business prerequisites in extensive data quickened ventures. Some examples are data

 DOI : 10.18231/2454-9150.2017.0029

warehouses work, transforming them into exact and express data definitions appropriate to additionally promotion and the task of data administration duties. They put a business data display in the middle – utilized end-to-end from investigation, outline, advancement, testing to data quality checks by data chairman. Likewise, they demonstrate that the strategy is legitimate past regular data warehouse stage.

Vikas S Shah et al. [17] have proposed an approach in consideration of multi faced compliance-aware data services (DSs) that enables the degree of distinction in business rules. They presented an approach to continuously monitor regulatory updates and rationalization to translate them into CSs. The research also presents runtime structure to evolve and govern compliance-aware DSs. In a relationship with business governs, the classification and the comparing execution of CSs into the DSs are recognized and inferred. Formulae to evaluate the level of refinement and appraisal criteria to screen administration of conveyed consistence mindful. DSs are represented with an illustration execution and approved based on quantity real sending cycles.

Xu Bihua *et al.* [18] have outlined a three layers engineering in light of Metadata which creates ETL process more compelling, multipurpose and compactable. The breaking data warehouse ETL devices were composed and executed through far reaching investigations of ETL hypothesis. The paper has inside and out concentrated the hypothesis of ETL apparatus which are firmly connected to the real task inside the data warehouse and proposed the arrangement of breaking data warehouse ETL device. The elements of metadata extricate, change and load to boring data warehouse was effectively accomplished. However, the proposition of ETL was obtained from the data Warehouse development, its application is more than inside data warehouse.

Jan H. Kroeze *et al.* [19] created and approved an IT appropriation structure to comfort associations with IT selection organization. To enhancing IT appropriation organization in associations as it tends to various worries amid IT reception the structure is valuable for partner purchase in, segregation, mistreatment and partner support. Before choosing appropriate methodologies for intercession the structure promoter for the need to comprehend the issue set. On that note, the properties of a structure to advance IT selection organization in associations must be extensive in nature to address its intricacy reception basic leadership in associations. In this way, there is no one to estimate fits-all approach in associations in IT selection organization.

Hsing *et al.* [20] have proposed an approach for a multipurpose medicinal database. The possibility of the framework is arranged. The initial segment of the examination is clarified in this paper. In Taiwan, the restorative related inquiries about that utilized the NHID database are evaluated. The data, the uses and also the techniques with the exception of in the inquiries about are incorporated into the examination. The discoveries are coordinate with the exploration incline around the world. After that examines the non-appearance of the present restorative database of Taiwan for data mining is talked about. The goal, the division, the approach and the data handle that were removed from the NHID database are compressed. Research discoveries presume that another composed database is in need to find concealed information from medicinal records.

Theodore Johnson *et al.* [21] have proposed a few data warehouses, which intertwine the advantages of old data warehouses, and data stream frameworks. The outer sources push embed just data streams into the warehouse with a colossal scope of inward entry times in the Existing model. A planning structure that handles the multifaceted nature protected by a stream warehouse was proposed. The structure contains see chains of importance, needs, data adaptability, inability to suspect updates, assorted variety of refresh occupations caused by various bury landing times and data quality between various sources, and transient over-burden. At last, they give an accumulation of refresh booking calculations and complete recreation analyses to delineate variables which influence their execution.

Abdolreza Hajmoosaei *et al.* [22] proposed a paper which researches the incorporated data warehouse, free data bazaar, subordinate data store, homogeneous circulated data warehouse and heterogeneous dispersed data warehouse. It has five unique data warehouse designs and a short time later an examination design will be depicted. The part of the data warehouse was clarified in first and a while later, a wide range of incorporated and dispersed models was depicted. Subsequently, the association's data handling framework will be perfect with its regularly developing data and quick on the grounds that the examination proposed circulated data warehouse engineering with high similarity with ideal design for associations.

To overcome the difficulties in existing works, we provide a new framework with trending technology. The proposed technology can increase the accuracy of the system. The following section elaborates the proposed work.

III. PROPOSED METHODOLOGY

The Support Vector Machine used to classify the data in our proposed work. Our developed system describes the Multipurpose Household Survey (MPHS) data and the Andhra Pradesh Government land record, and it can be classified by using SVM Classifier. After that the classified information can be saved in the mongo database for the purpose of secure storage. Mapping Co-ordinator provides access rights to the user. The fuzzy based ranking function can be used to retrieve the user need information for efficient retrieval. The architecture of our framework is given below Fig. 1.

 DOI : 10.18231/2454-9150.2017.0029

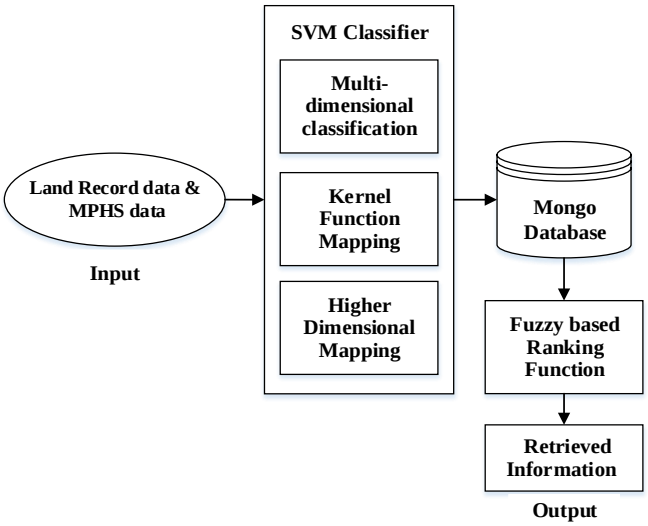

Figure 1: Architecture of our proposed system

The Multipurpose Household Survey (MPHS) data and the land records data of the Andhra Pradesh Government can be gathered in our proposed architecture that act as the input of the proposed system. By using SVM classifier the collected land record data can be classified. The performance in the support vector machine is the classification using the different process to give the efficient results. Then the classified information is securely stored in the mongo database. After that the fuzzy ranking function effectively retrieved information. The fuzzy ranking function using ranking format depends on the fuzzy rules. It provides efficient retrieval of information to the user.

Data collection of land records, Classification of the information, database storage, and Information retrieval are the main modules of the proposed system. In Fig. 2 Show that the work flow of our framework.

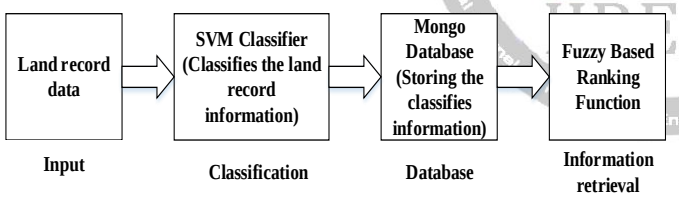

Figure 2: Block diagram of the proposed system

3.1 Data collection for the Land Record Information

Main process of our proposed work is Data collection. The Multipurpose Household Survey (MPHS) data is used to achieve better data collection and land records data is collected from different sources. The land record departments and settlement Commissioner collected the state level land records. The District Inspector as well as Taluk Inspector gathered the information of particular District land records. A function of the Land Records department is described in the following paragraphs.

1. By creating field operations the up-to-date survey, classification and settlement records are collected and protected. It allows updating the changes in survey records.

2. By utilizing Analytical information, all data dependent on land records are collected and maintained effectively.

3. For preserving survey and target land records are the dispute in civil courts. It simplifies the methods and reduces the cost.

4. Record of Rights can be supervised by periodical monitoring, maintenance, and reconstruction of the boundary marks of unique fields.

5. Compensation operations are performed to achieve periodical enhancement.

6. Appropriate preservation and pervasive scale are organized to obtain village site surveys.

7. Up-to-date changes are performed in village survey to manage the records.

8. All tahsil maps are reprinted and separated under various departments. It provides easy access to records to the user.

9. Finally, the concatenations of the revenue officers in survey and settlement matters are taking place.

The information present in the records is converted into uniform format. Then the collected information used as input for the classifier to classifies the information.

3.2 Classification

The Support Vector Machine (SVM) are used in our proposed system to classify the information. During the data collection, distinct types of data to be collected but the information is not in proper order. So the user required information is not be accessed easily. Both classification and regression challenges SVM is a supervised machine learning algorithm which was the benefit. We plot each data item as a co-ordinate in N-dimensional space where, n is the number of aspects of the information in this algorithm. That contains the value of each feature being the value of a distinct coordinate. Then, we perform the classification process by determining the hyper-plane that differentiates the land records information. Support Vectors are easily co-ordinates of individual consideration. In the presented methodology, SVM classification includes different process of classification is non-linear classification, kernel function mapping, and higher dimensional mapping as given below.

a. Non Linear Classification

In a classification process, the non-linear classification is the easiest way of separate the two groups of data such as a straight line (1D), the flat plane (2D) or an N-dimensional hyper plane. The basic diagram of non-linear classification is shown in Fig. 3. A non-linear function is educated with a linear learning machine in a high-dimensional feature space. While the quantity of the system is disciplined by a restriction that does not depend on the depth of space. So the distinct types of information can be classified in the feature space. After that, the classified data transform into the kernel function mapping to map the desired data in-between them.

 DOI : 10.18231/2454-9150.2017.0029

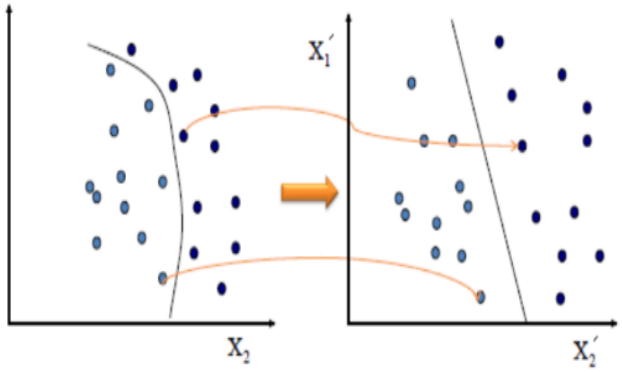

Figure 3: Non-linear Classification

b. Kernel Function Mapping

In this framework kernel function mapping is another process of SVM classification for separating for user convenient. In nonlinear framework Kernels make linear models performance. It displays linear patterns by mapping data to higher dimension. Then, put on the linear model in the new input space

Mapping ≡ changing the feature representation

Such mappings can be valuable to measure in general but kernels give such mappings for free. We need a nonlinear dividing line that shown in Fig. 4 in this situation. SVM maintains this by utilizing a kernel function to map the data into a different space rather than proper nonlinear curves to the data where a hyper plane can be utilized to do the partition. Depending on the dividing line the data can be easily partitioned. Also kernel mapping performs the mapping function for relevant data combined within the related information in the field.

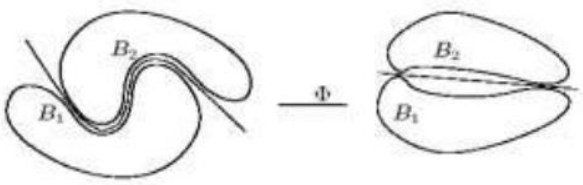

Figure 4: Kernel Function Mapping

After that, for the efficient result, the kernel function may transform the data into a higher dimensional space to make it possible to perform the mapping.

c. Higher Dimensional Mapping

In our proposed work we are using higher dimensional mapping for mapping the data from lower dimensional to higher dimensional. Higher dimensional mapping is the finite dimensional one that provides effective separation of data, but the transformation as applied to the positive input. In this work, for transform, the data into higher dimensional space, the kernel function mapping performs the implicit conversion. Already in the original space, the data may be linearly separable.

Hence, our proposed work classifies the input data using non-linear classification to give the separated field of information. After that perform the kernel function mapping ignores the irrelevant data and maps the relevant data. Then the relevant data in the lower dimensional can be sent to a higher dimensional mapping function that gives the clear information about the land record information. Hence, the SVM classifier classifies the different land record in particular order. The individual personal details can be classified in separate order.

3.3. Mongo database

Mongo database stores the classified data from SVM classifier. Also, it is the document dependent database. So the separable data can be stored in each field of the database for user convenient retrieval. The Mongo database is a database that stores huge objects or files. The database has a lot of advantages for storing the files in different fields. So the information from the desired field of the database can be accessed by the user.

3.3.1. Mongo DB Data Model

Mongo Database is a document database. Each database contains collections of documents. Each document can be different with the varying number of fields. The size and content of each document can be different from each other. Benefit of the Mongo Database is a shot-platform, the document oriented record that gives high level performance and availability also it provides better scalability.

3.3.2 Data as Documents

Mongo DB records contribute to have a number of the document in a single database because in a relational database information is a normally extension across many tables. In each blog document that contains multiple comments, multiple tags, and multiple sections, each expressed as fixed array that shown in fig 3.3.2. In this work based on the particular format the land record information is stored in the database. That includes the user name, desired land information, the location of the land, length of the land, amount of land stored in each field of the database.

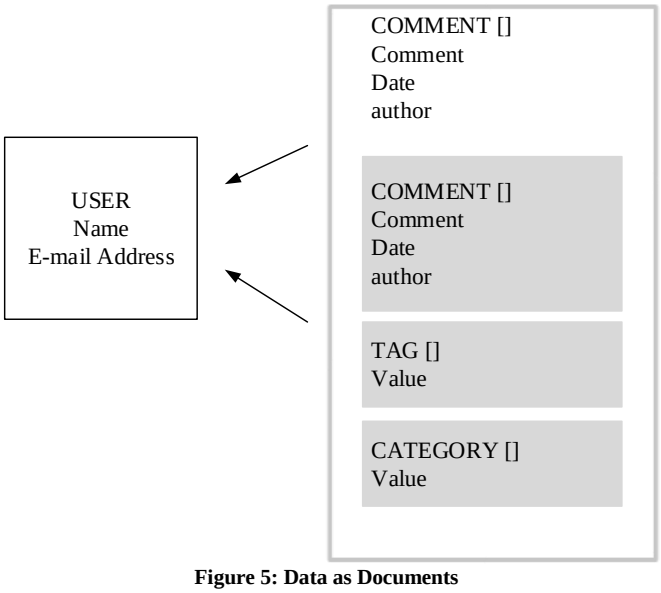

Figure 5: Data as Documents

 DOI : 10.18231/2454-9150.2017.0029

Data as documents is uncomplicated for developers, faster for users. In our structure, the classified data can be stored in the mongo database as rows and columns also Individual information stored in each field in the database. So the information can be retrieved from the mongo database in particular format. The user need information will be retrieved from the database using fuzzy based ranking function techniques that briefly described as given the sub section.

3.3.3 Mapping Co-ordinator

The mongo database include mapping co-ordinator for storing and retrieving the data in an effective way. For maintaining and implementing standards for hardware and software to assure compatible, reliable and user-friendly operation the mapping coordinator is used. Also, it handles upload/download, updates of information for the user convenient. In major usage of the mapping co-ordinator retrieved the information from the data base in the secured manner.

3.3 Information Retrieval

In our proposed system, a novel Fuzzy Based Ranking Function is presented in our proposed system. The intuition behind using fuzzy logic is that it provides acceptable way to retrieve the information using fuzzy logic rules. The fuzzy logic based ranking function could be easily displayed, extended and retrieved. Ranking functions recovered a ranking value for all rows on a separation and it is non-deterministic. The ranking function is the method of ordering the fuzzy numbers for data. In this work the ranking function provides the rank for individual field in the database. So the user provides request data can be accessed depending on the ranking function.

3.3.1 Fuzzy rule base

The most familiar way, the fuzzy rule described the knowledge convert into a natural expression that is represented as,

IF the premise (antecedent) THEN conclusion (consequent)
IF the premise1 (antecedent) AND premise2 (antecedent) THEN conclusion (consequent)

The above rule commonly referred as IF-THEN rule format. It consistently conveys an inference such that if we know a fact (premise, hypothesis, antecedent), then we can infer, or derive, another fact called a conclusion (consequent). Another form in multiple conjunctive antecedents is referred as IF–THEN fuzzy rule-based form. In this method, fuzzy rules are taken from the familiar knowledge of Information Retrieval system. In our proposed system, the using fuzzy ranking function for retrieving the clear information about the land record. The same name of the field will differ because each field of information contains the different ranking value. So the user required information is accessed easily.

IV. EXPERIMENTAL RESULT

We presented the evaluation of our proposed system in this section. In our experiments, our aim is to classify the different types of data and the user can retrieve the information about land records in Andhra Pradesh government in anywhere. We proposed the classification technique provides the better performance to classify the land records of Andhra Pradesh government.

The input of the classification process is shown in Fig. 6 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 SVM 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 SVM classifier for user convenient. That different function includes Non-Linear Classification, Kernel Function Mapping, and Higher Dimensional Mapping. In Non-linear classification separates the different district name and the land holders name and their properties. After that the kernel mapping using the dividing line to partition the separate fields of the data, but that are placed in lower dimensional mapping. So using Higher Dimensional Mapping for each of the district and land holder information and their properties to be separated and drive the efficient result in a higher dimensional space.

no	region	name	mailid	property	district
1	Guntur	Rajeev eddy	rajeev@gmail.com	0 acre0yard s2 0 feet	1
2	Narsapur	kunkana patel	kunkadupti@gmail.com	0 acre0yard s2 0 feet	1
3	Chegunta	radhakrishnan	rachuruvidya@gmail.com	0 acre0yard s2 0 feet	1
4	Dubbak	lekshmidevamma	Lakshmidev11@gmail.com	0 acre0yard s2 0 feet	2
5	Guntur	Anji reddy	anji12221@gmail.com	0 acre0yard s2 0 feet	2

Figure 6: Input of the classification process

 DOI : 10.18231/2454-9150.2017.0029

sl no	region	name	mailid	property	District
1	Guntur	Anji reddy	anji12221@gmail.com	0 acre0yard s2 0 feet	2
2	Guntur	Rajeev eddy	rajeev@gmail.com	0 acre0yard s2 0 feet	1
3	Chegunta	radhakrishnan	rachuruvidya@gmail.com	0 acre0yard s2 0 feet	1
4	Dubbak	lekshmidevamma	Lakshmidev11@gmail.com	0 acre0yard s2 0 feet	2
5	Narsapur	kunkanapate	kunkadupti@gmail.com	0 acre0yard s2 0 feet	1

Figure 7: Classification output

After performed the classification process the field of the data will be separated in the different label. That are shown in below Fig. 7. During the development process Mongo database provides flexibility. The Mongo database can merge the fields of data that can achieve user requirements. In our proposed system we are using mongo data base for storing different kinds of user details. That detail describes the land holder of the Andhra Pradesh government. The mongo database provides secured storage, flexibility of retrieval and easy to access.

The user required information is retrieved from the database or other resources is described in information retrieval. The information is retrieved from the mongo database using the fuzzy ranking function in our system. Hence the fuzzy-based ranking function using the weighting scheme to differentiate the record data and retrieve the desired information for the user. For example, any user request the particular information about the district or land holder means they retrieve the related field of information that is given below.

V. PERFORMANCE ANALYSIS

In this section, performance analysis between different classifiers and the database and that compared with our system. Our proposed system using SVM classifier classifies the different information about land records of Andhra Pradesh government. After that for efficient storage the classified information is stored in the mongo data base [23]. The efficiently storage and retrieval of relevant information is the main usage of the mongo data base. [24], [25]. Then the classified information is retrieved by using fuzzy based ranking function [26]-[28].

The proposed system using SVM Classification that compared with other different types of classification that shown in Fig. 8 to 11. The result of the comparison is given in the table. The proposed classification better than other classification that using different parameters are efficiency, accuracy, scalability that are shown in below. The efficiency means the classification given the better performance compared with other classification. The accuracy of the result in SVM classification is very high. Kappa statistics, Mean absolute error, Root mean square error, relative absolute error, root relative squared error is the accuracy parameters.

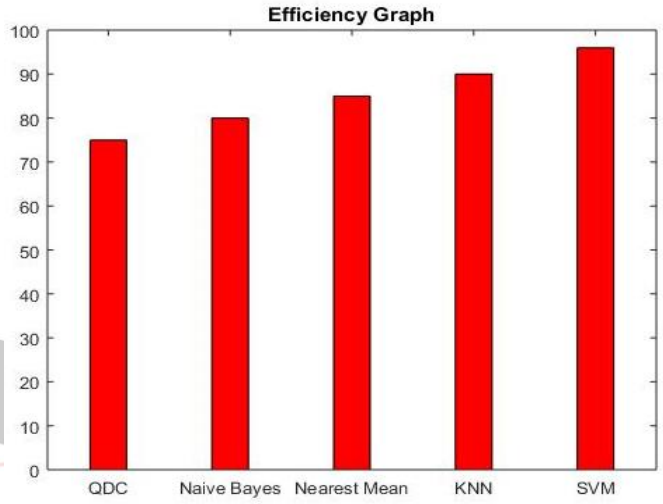

Figure 8: Efficiency of SVM Compared with other classifiers

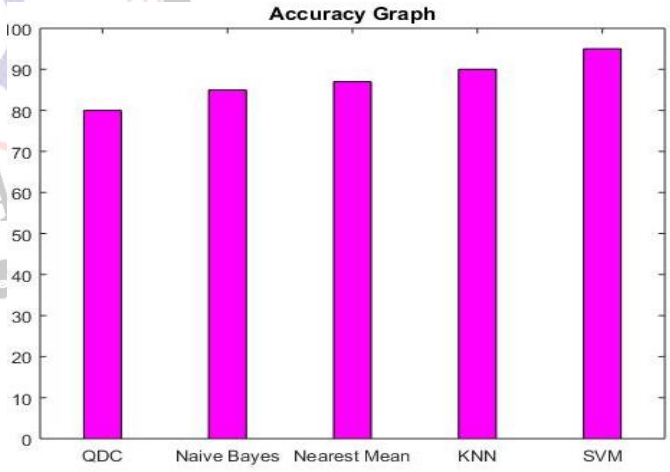

Figure 9: Accuracy of SVM compared with other classifiers

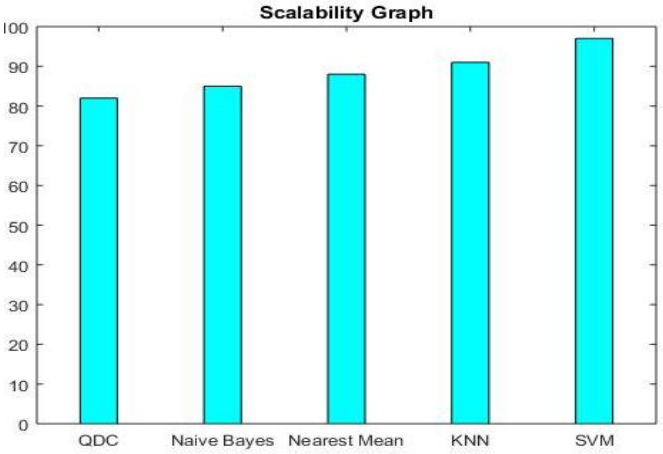

Figure 10: Scalability of SVM Compared with other classifiers

 DOI : 10.18231/2454-9150.2017.0029

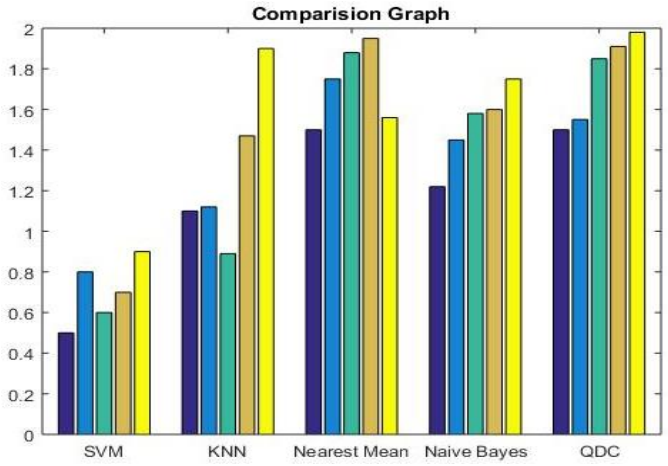

Figure 11: Accuracy parameter compared with SVM and other classifiers

Table 1: Classifier Performance

Parameters	SVM	KNN	NN	Naïve bayes	QDC
Kappa statistics	0.5	1.1	1.5	1.22	1.50
Mean absolute error	0.8	1.2	1.15	1.45	1.55
Root mean squared error	0.6	0.89	1.88	1.58	1.85
Relative absolute error	0.7	1.47	1.95	1.60	1.91
Root relative squared error	0.9	1.90	1.56	1.75	1.98

Various accuracy measures for various classifier models used to predict data from given record information are presented in the above table [29]. The best classifier model is found out using the various criteria used for evaluation. The kappa statistics, mean absolute error, root mean squared error, relative absolute error and root relative square error are the criteria used for evaluation with different classifiers. In our system using the mongo database is compared with MySQL database [30] [31]. The result of the table described the mongo database very efficient than the MySQL database. The fig shows that amount of time the database storing the data and retrieving the data in efficient manner compared with the MySQL database.

V. CONCLUSION

This paper is proposed to achieve the great accuracy. The chief importance of this paper is to provide easy access and retrieval of land related information to the common user. In our proposed work, land information is gathered from resources, then Support Vector Machine classifies the information. Mongo DB stores all information and Mapping Co-ordinator helps to provide access rights and security. Finally, fuzzy based ranking function utilized to retrieve the relevant information. The mongo database can store a large quantity of data. It does not have any restriction or limitations. This proposed work allows decision makers and policy planners to access the relevant data (current data). It improves the accuracy of the system. SVM classifies helps to classify the content. So, it avoids the conflict while accessing the data. Mapping co-ordinator provides high security to the user. It allows authority to upload/download the new data. It establishes a restriction on access to restricted information.

REFERENCES

[1] March, Salvatore T., and Alan R. Hevner, "*Integrated decision support systems: A data warehousing perspective*", Decision Support Systems 43, no. 3, pp. 1031-1043, 2007.

[2] Gupta, Himanshu, and Inderpal Singh Mumick, "*Selection of views to materialize in a data warehouse*", IEEE Transactions on Knowledge and Data Engineering 17, no. 1, pp. 24-43, 2006.

[3] Ariyachandra, Thilini, and Hugh Watson, "*Key organizational factors in data warehouse architecture selection*", Decision support systems 49, no. 2, pp. 200-212, 2010.

[4] Araque, Francisco, Alberto Salguero, and Maria Abad, "*Application of data warehouse and Decision Support System in soaring site recommendation*", Information and Communication Technologies in Tourism, pp. 308-319, 2006.

[6] Smith, Matthew, Christian Szongott, Benjamin Henne, and Gabriele von Voigt, "*Big data privacy issues in public social media*." 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST), pp. 1-6, IEEE, 2012.

[7] 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.

[8] Katharina Ebner, ThiloBühnen, Nils Urbach, "*Think Big with Big Data: Identifying Suitable Big Data Strategies in Corporate Environments*", IEEE, pp. 978-1-4799-2504, 2014.

[9] Bello-Orgaz, Gema, Jason J. Jung, and David Camacho, "*Social big data: Recent achievements and new challenges*", Information Fusion 28, pp. 45-59, 2016.

[11] Pan, Gang, Guande Qi, Zhaohui Wu, Daqing Zhang, and Shijian Li, "*Land-use classification using taxi GPS traces*", IEEE Transactions on Intelligent Transportation Systems 14, no. 1, pp. 113-123, 2013.

[12] Katkar, Vijay, Surupendu Prakash Gangopadhyay, SagarRathod, and Aakash Shetty, "*Sales forecasting using data warehouse and Naïve Bayesian classifier*", Pervasive Computing (ICPC), 2015 International Conference on, pp. 1-6. IEEE, 2015.

[13] Chou, Jui-Sheng, Min-Yuan Cheng, Yu-Wei Wu, and Anh-Duc Pham, "*Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute*

 DOI : 10.18231/2454-9150.2017.0029

classification", Expert Systems with Applications 41, no. 8, pp. 3955-3964, 2014.

[14] Chou, Jui-Sheng, Min-Yuan Cheng, Yu-Wei Wu, and Anh-Duc Pham, "*Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification*", Expert Systems with Applications 41, no. 8, pp. 3955-3964, 2014.

[15] Demirkan, Haluk, and DursunDelen, "*Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud*", Decision Support Systems 55, no. 1, pp. 412-421, 2013.

[16] TorstenPriebe, 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.

[17] Vikas S Shah, "*An Integrative Methodology to Enable Compliance aware Data Services Accelerating Consumer Regulatory Governance Paradigms of an Enterprise*", 13th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2016.

[18] Li Jian, Xu Bihua, "*ETL Tool Research and Implementation Based on Drilling Data Warehouse*", Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010), 2010.

[19] OsdenJokonya, Jan H. Kroeze, John A. van der Poll, "*A Framework to Assist Organizations with IT Adoption Governance*", IEEE 6th International Conference on Cloud Computing Technology and Science, 2014.

[20] Wang, Hsing-I, "*The Preliminary Survey for the Design of Multipurpose Medical Databases for Taiwan – the Usage Oriented Approach*", Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2011.

[21] Lukasz Golab, Theodore Johnson, and VladislavShkapenyuk, "*Scalable Scheduling of Updates in Streaming Data Warehouses*", IEEE Transactions on Knowledge And Data Engineering, 2012.

[22] Mehdi Kashfi, AbdolrezaHajmoosaei, "*Optimal Distributed Data Warehouse System Architecture*", IEEE Fourth International Conference on Big Data and Cloud Computing, 2014.

[23] Horng, Shi-Jinn, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui-Lin Lai, and Citra Dwi Perkasa, "*A novel intrusion detection system based on hierarchical clustering and support vector machines*", Expert systems with Applications 38, no. 1, pp. 306-313, 2011.

[24] Fan, Weiguo, Michael D. Gordon, and Praveen Pathak, "*A generic ranking function discovery framework by genetic programming for information retrieval*", Information Processing & Management 40, no. 4, pp. 587-602, 2004.

[25] Zhang, Hao, Alexander C. Berg, Michael Maire, and Jitendra Malik, "*SVM-KNN: Discriminative nearest neighbor classification for visual category recognition*", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 2, pp. 2126-2136. IEEE, 2006.

[26] Cadenas, J. M., and J. L. Verdegay, "*Using ranking functions in multiobjective fuzzy linear programming*", Fuzzy sets and systems 111, no. 1, pp. 47-53, 2000.

[27] Hu, Qinghua, Daren Yu, and ZongxiaXie, "*Information-preserving hybrid data reduction based on fuzzy-rough techniques*", Pattern recognition letters 27, no. 5, pp. 414-423, 2006.

[28] Su, Ja-Hwung, Bo-Wen Wang, Tien-Yu Hsu, Chien-Li Chou, and Vincent S. Tseng, "*Multi-modal image retrieval by integrating web image annotation, concept matching and fuzzy ranking techniques*", International Journal of Fuzzy Systems 12, no. 2, pp. 136-149, 2010.

[29] Yang, He, Qian Du, and Genshe Chen, "*Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification*", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5, no. 2, pp. 544-554, 2012.

[30] Saikia, Amlanjyoti, Sherin Joy, and Roseline Mary Dhondup Dolma. "*R, Comparative Performance Analysis of MySQL and SQL Server Relational Database Management Systems in Windows Environment*." (2015).

[31] Shute, Jeff, RadekVingralek, Bart Samwel, Ben Handy, Chad Whipkey, Eric Rollins, MirceaOancea et al. "*F1: A distributed SQL database that scales." Proceedings of the VLDB Endowment* 6, no. 11 (2013): 1068-1079.

 DOI : 10.18231/2454-9150.2017.0029