

# ROLE OF DATA MINING IN E-GOVERNMENT FRAMEWORK

Reem Razzaq Abdul Hussein<sup>1</sup>

<sup>1</sup>informatics Institute of Postgraduate studies ,  
Baghdad, Iraq  
reemrazzak@yahoo.com

Dr. Salih Mahdi Al-Qaraawi<sup>2</sup>

<sup>2</sup>Computer Engineering Department,  
University of Technology, Baghdad, Iraq  
drsalihalqaraawi@gmail.com

Dr. Muayad Sadik Croock<sup>3</sup>

<sup>3</sup> Computer Engineering Department,  
University of Technology, Baghdad, Iraq  
University  
muayadkrook@yahoo.com

**Abstract** - In e-government, the mining techniques are considered as a procedure for extracting data from the related web application to be converted into useful knowledge. In addition, there are different methods of mining that can be applied to different government data. The significant ideas behind this paper are to produce a comprehensive study amongst the previous research work in improving the speed of queries to access the database and obtaining specific predictions. The provided study compares data mining methods, database management, and types of data. Moreover, a proposed model is introduced to put these different methods together for improving the online applications. These applications produce the ability to retrieve the information, matching keywords, indexing database, and performing the prediction from a vast amount of data.

**Keywords** - E-government, data mining, database management, prediction.

## I. INTRODUCTION

Recently, the government services have been produced to facilitate the life requirements. Moreover, it transform the services from traditional into electronic copy in two major paths: information technology and communication (ITCO) as well as the internet [1]. The internet plays a major role in e-government platforms to proliferate the e-services. The information is gathered from different fields of e-government, such health filed, detection of criminals data, education, and t data related to smart\_cards [2]. The database is the backbone of constructing the e-government infrastructure for offering better accessing for demanded data and managing the services delivery to people. It is important to note that the database faces many challenges with incremental data, such as difficult to analysis, recognizing, interpreting vast data. These problem had been solved by data mining technologies that are able to extract the new knowledge, retrieve on-demand, and prediction to make a decision [3], [4].

The mining, sometimes named Knowledge Discovery in Databases (KDD), can be defined as the process to extract a knowledge then analyzing it. The first major task of mining is the classification that is used for classifying objects by category value (label or goal). As a result, each object belongs to a specific class, such as decision tree which is common one [5]. Many classification algorithms have been suggested for managing the data, such as Quinlan, CHART, FACT. The other

classification algorithm, Naïve Bayes, deals with an independent variable and k-nearest-neighbor depending on the distance of parameters, etc. The main idea of classification methods is to obtain the greatest amount of accuracy to predicting the right class for each case in the data [6].

The clustering, which is the second task of mining methods, is known as unsupervised learning. It is a set of entities are similarly grouped in a cluster and different from the other clusters. In summary, both classification and clustering methods can be used to analyze the data government, and then passing the analyzed data to be used for measurements as a loop, see in Figure (1).

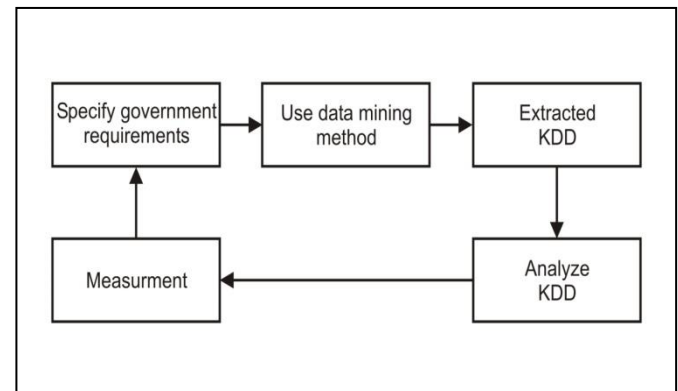


Fig. 1. Data mining looping

The mining methods can be more effective when they are combined with machine learning, statistic method, database management, and artificial intelligence for improving results.

## II. RELATED WORK

Numerous research works have been produced to improve the data management. The related works focusing particularly on data mining methods and arising issues as follows :

In [7], the authors described an implementation framework that constructs a prototype for similarity search. It included distributed database management, and load balance technique. They presented the statistics of automatic performance and relayed on indexing technique. The prototype was constructed

on a modular design that consisted of many modules with different specific jobs. The presented prototype faced some problems that need to develop a platform to get an accurate result.

In [8], the authors proposed a model of DNA based on decision tree indexing method. Various of indexing method types have been described. The study depended on proteins sequence data set as well as proposing a model to find an efficient way to retrieve, search, insertion, deletion elements, minimize space, query size. The complex rules had been generated with redundant data to satisfy the memory bottleneck that required to fit all rules in memory. The result of the experiment was compared with another algorithm indexing method.

In [9], the author proposed a model of national identification that relays on the distributed database of smart cards. The data of smart cards included identification information, which consisted of the biometric, attribute, and some personal information.

In [10], the authors proposed searching system depending on keywords. They worked to improve the performance of searching task and efficiently manage the database. The authors depended on mining algorithm, named FP\_Growth indexing database, and also the clustering technique of graph and schema methods.

In [2], the authors proposed and discussed the effectiveness of data mining method on the massive amount of data that was generated from online e-services using java programming language. They proposed a model for grievance redressal to be included in e-government framework. The proposed model computed the cost of efficiency for learning different applications such as classification data, detection, and tracking that had been solved in mining technique.

In [11], the authors presented a prototype for optimizing data mining in e-government, emphasizes retrieve information from a database. The modern technique used cloud computing, and MySQL data indexing databases management.

### III. THE INDEXING WITH DATA MINING TECHNIQUE

Indexing method, related to of data structure, has a major role in improving the speed of data retrieval from the database. It is used to avoid the whole search in every tuple in database tables each time. There are different approaches for matching and building databases purposes with full consideration to storage of e-government services in the long term. Mining techniques can be overlapped with index methods [12]. The index approach is divided into two main categories: exhaustive, and index search [13][16]. The index search is classified into:

1. R+tree, Btree method that inverts files including a reference to document for each word, bitmap. It is the type of indexing that depend on logical operations to perform queries and stores its data as bit index (0,1) [17].

2. Sequence method depends on decision tree indexing method.
3. The hybrid method is a mix of different indexing methods.

The necessary resources need to manage databases for an insignificant amount of efficient retrieval data when needed. However, databases do not have the ability to make decisions or extraction an information. Hence, the importance of data mining is to discover knowledge from databases [18]. The primary stages of data mining are described as follows:

### IV. THE TYPES OF DATA IN DATABASE

1. There are different types of data that can be dealing with mining methods. These types can be summarized as follows:
2. Flat files: is one of the most common sources of mining data, especially in the field of research. This kind of files may be text or binary. It is the simplest kind of data mining method, and it can be used time series or transactions, etc.
3. Relational: the database structure is relational, it means that it consists of a group of tables, which are related by a unique key. Each table consists of records belong to items, column or some time called attribute.
4. Transaction: a transaction database is a set of data raw representing transactions, each of which consists of timestamp identifier and a group of items with a single timestamp.
5. Multimedia: this type of databases include different types of data, such as video, audio, text media and, and images.
6. Object-oriented databases: this type of data is high of dimensions that face different challenges to be solved using many methods of mining.
7. Time-Series databases: Time-series databases consider time concern with data, such as market stock data or logged activities data. These databases are iteratively changed with time when new data have been added. In real-time, the main challenge faced by researchers is the analysis, which can be another trend in mining algorithm.
8. Spatial database: spatial databases store traditional data in addition to spatial data, including local and global positioning, as well as maps. The spatial databases face main challenges, particularly the data mining algorithms in terms of classification [20].
9. Amongst the research work so far, there are different methods in data indexing, each of which is applied to the specific case and some researchers use hybrid strategy. The previous researches, focused on database management, have considered the types of data, data source, data mining methods as shown in Table (1). There are different data sources in a different area in e-government, for example, Khalid Jaber, and et.al in [8] created index database depending on mining using C++, SQL software based on protein sequence. In addition, C'ecile Favre, and et.al in [21] and Peter Christen in [18] used the text data type, while Yasir Safeer and et.al in [22]

used flat data and Weijie Kong , and et.al in [23] employed spatial databases. Moreover, Abdul Khanadar in [24] utilized spatial database with the is smart card source of data, while Anca Mehedinu [13] and Xiufeng Li in [26] did not attempt mining approaches.

TABLE (1): SUMMARY OF DATABASE AND DATA MINING APPROACHES

Authors	DBMS Description	Dataset	mining Approach	Remarks
Khalid Jaber, and et.al	Relational DB my SQL C++	Protein sequence	Yes	Reduced the searching space using Decision Tree index.
Aehedinu	Relational DB	/	No	Btrees, hash, and bitmap indexing, in order to optimize queries.
C'ecile Favre and et.al	Oracle	Titanic	Yes	Decision-tree-based and Bitmap indices methods.
Peter Christen	DB my SQL text	/	Yes	Select a name matching technique using a clustering method.
Yasir Safeer, and et.al	Text Miner Flat data	24 document	Yes	Text data mining.
Weijie Kong	Spatial Databases	GIS	Yes	Classification and clustering 2017
Ihsan A. Kareem, and et.al [24]	/	bank, diabetes, credit approval data	Yes	C4.5 Decision Tree
Abdul Khanadar, and et.al	spatial data	Data smart card	Yes	Clustering 2016
Xiufeng Liul and et.al	high dimensionality / PostgreSQL	/	No	Data Warehouse for e-government 2010

**V. THE PROPOSED MODEL**

This paper suggests a model, derived from previous works that use data mining algorithms with some developments, for two approaches. The first one is for searching a pattern over a large amount of useful data within the field of e-government, using specific keywords in an advanced way. The second approach is to predicting and analyzing the data for enabling a decision making to overcome the difficulties of analyzing. The

related work knowledge has been improved in a proposed model including the following:

1. Working online through a web application that can be accessed anywhere, any time.
2. The mining technique of classification can be decision tree, Navie bays, etc.
3. Applying index technique as mention earlier.

The summery stages are described as a flowchart, as shown in Figure (2):

1. Collecting the data of e-government in Iraq: In the first step, the real data (dataset) has been collected from the smart card organization.
2. Preprocessing: In this step of the proposed model, the outlier and redundant data have been removed and cleaned.
3. Apply Classification algorithm: This stage includes construction decision tree, that requires indicating the split value at the beginning, then selecting the highest entropy of attribute and place on the root of tree permit one branch to each value. The step(1),(2)(3) repeats until each leaf node(decision node ) there no split process when each attribute belonged into a specific class, at this step, we can employ mining technique for indexing database and prediction.
4. Employing a data mining method to a predicate and reflect the design of database index depending on the adopted dataset.
5. Applying optimization method using evolutionary strategy.
6. Searching in the queries tuple includes multiple keyword searches.
7. Results and reports app ear through a web application.

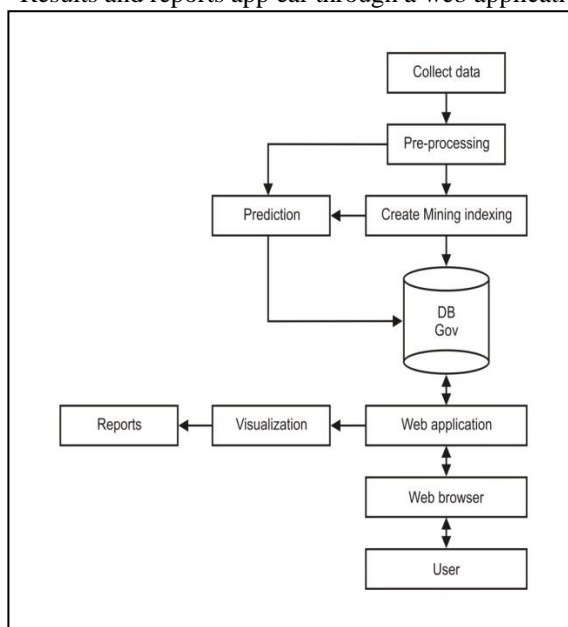


Fig. 2. The proposed flowchart

## VI CONCLUSION

E-government has improved the way of interaction between the government and the people by creating a model that adopts data mining techniques instead of traditional techniques. The mining model dealt with quality data, amount of data and online transparent access to e-services. In this paper, a survey about e-government and mining technique has been introduced. In addition, the rising problems and recommendations, extracted from the previous works, between data mining and data management systems have been addressed. Finally, we proposed a model that depends on advantages of indexing theories that overlap mining theories with database management system. This is to read fewer data with more accuracy and faster access method

## REFERENCES

- [1] GK Rao, "Decision Support For E-Governance: A Text Mining Approach", International Journal of Managing Information Technology (IJMIT), Vol.3, No.3, August 2011.
- [2] Sangeetha G. and Rao L.M., "Modeling of EGovernance Framework for Mining Knowledge from Massive Grievance Redressal Data". International Journal of Electrical and Computer Engineering (IJECE), 6(1), 367-374, 2016.
- [3] V.J.Singh, A. Change, "Evolving E-Governance through Cloud Computing based environment", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 4, 2014.
- [4] Chen, H., "Digital Government: technologies and practices; 2003. Decision Support Systems", vol 34(3), pp 223-227, 2003.
- [5] Yi Yang, Wenguang Chen, "Taiga: performance Optimization of the C4.5 Decision Tree Construction Algorithm" ISSN 11007-0214 06/11 ,pp415-425, Volume 21, Number 4, August 2016.
- [6] Tongwei yuan, Peng Chen "Data Mining Applications in E-Government Information Security", Procedia Engineering vol 29 pp235 – 240, 2012.
- [7] M. Batko, D. Novak, and P. Zezula, "MESSIF: Metric similarity search implementation framework," in First International DELOS Conference, Pisa, Italy, Revised Selected Papers, ser. Lecture Notes in Computer Science, vol. 4877. Springer, pp. 1-10, 2007.
- [8] Khalid Jaber, Rosni Abdullah, and Nur Aini Abdul Ra, 'Adapting decision tree-based method to index large DNA-protein sequence datasets'. The Research Bulletin of Jordan ACM. 2, 3, 57-73, 2011.
- [9] Olatubosun Olabode, "Smart Card Identification Management Over A Distributed Database Model ", Journal of Computer Science 7 (12): 1770-1777, 2011 ISSN 1549-3636.
- [10] Pradeep M. Ghige, Ruhil R. Kabra, "Relational Keyword Search System", International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November, 2014.
- [11] Sangeetha G, "ODM: Modelling Optimization of Datamining in e-governance Framework", International Journal of Engineering Research and General Science, ISSN 2091-2730, Volume 6, Issue 1, January-February, 2018.
- [12] D. Kettani, "E-Government Applications in the African Context", Tangier Morocco, 2014.
- [13] Anca Mehedintu, "Indexing Strategies For Optimizing Queries On MySQL", Annals of the University of Petroşani, Economics, vol. 10, issue 4, 201-214, 2010.
- [14] José María Luna; Alberto Cano; Mykola Pechenizkiy; Sebastián Ventura, eds, "Speeding-Up Association Rule Mining With Inverted Index Compression", JARVIE e-ISSN: 2395-4396. Volume/Issue: Volume 3 Issue 2 2017.
- [15] Khalid Mohammad Jaber, Rosni Abdullah, and Nur'Aini Abdul Rashid. Adapting and Enhancing the Searching Algorithm Based on Decision Tree Indexing for Large DNA-Protein Datasets. The 2011 Conference on Innovations in Computing and Engineering Machinery (CICEM'11)- ACM Chapter Conference, Vol. 2, pp. 113-122, Grand Hyatt Amman Hotel, Amman, Jordan, 5-6 September 2011.
- [16] Amar Jeet Singh, "E-Government database: A Retrospective study" Indian Journal of Computer Science and Engineering Vol 1 No 2, 66-73, 2010
- [17] Shu-Hsien Liao, Pei-Hui Chu, Pei-Yuan Brief Survey of Text Mining: Classification Clustering and extraction techniques", KDD Bindas, August 2017.
- [18] Peter Christen, "A Comparison of Personal Name Matching: Techniques and Practical Issues, 2006.
- [19] Ms. Aruna J. Chamatkar, Dr. P.K. Butey, "Importance of Data Mining with Different Types of Data Applications and Challenging Areas", ISSN: 2248-9622, Vol. 4, Issue 5( Version 3), pp.38-41, May 2014.
- [20] Ramez Elmasri, "Database Systems models, Languages, Design, and application programming, six edition, ISBN10-0-13-214498-0, 2011.
- [21] C'ecile Favre and Fadila Bentayeb, "Bitmap index-based decision trees", International Symposium on Methodologies for Intelligent Systems ISMIS: Foundations of Intelligent Systems, Springer, pp 65-73, 2005.
- [22] Yasir Safeer, Atika Mustafa, Anis Noor Ali" Clustering Unstructured Data (Flat Files), IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 2, 2010.
- [23] Weijie Kongl Harbin, "Analyzing the Usage of Data Mining in Spatial Database" International Journal of Internet of Things and its Applications Vol. 1, No. 1 (2017), pp. 1 <http://dx.doi.org/10.21742/ijiota.2017.1.1.01>
- [24] Ihsan A. Kareem, Mehdi G. Duhaime, "Improved Accuracy for Decision Tree Algorithm Based on Unsupervised Discretization, IJCSMC, Vol. 3, Issue. 6, June, pg.176-183, 2014.
- [25] Abdul Khanadar, Bhanupriya Sharma, Shambhavi Srivastava, "Data Mining from Smart Card Data using Data Clustering", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 1 (2016) pp 347-352.
- [26] Liu, Xiufeng; Luo, Xiaofeng, "A Data Warehouse Solution for e-Government", International Journal of Research and Reviews in Applied Sciences, 4(1), 101-105., 2010.