IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol-6, Special Issue-4, May 2018 E-ISSN: 2347-2693

Change Detection Analytics on Water Contamination using Decision Tree based Classification

P. Shanmugavadivu[1*], P. Kavitha[2], S. Dhamodharan[3]

[1,2,3] Department. of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University), Dindigul, India

[*]*Corresponding Author: psvadivu67@gmail.com , Tel.: +91-94437-36780*

Available online at: www.ijcseonline.org

***Abstract*—** The exponential growth as well as the availability of data has triggered the development of data analytics tools to harness the power of information hidden in data. These advancements have grabbed the attention of the researchers across the globe. The research frontiers of data analytics is expanding so rapidly, covering many fields including business intelligence, finance, market analysis, science, environmental studies, resource management, weather forecasting and outlier analysis. This paper describes the design of a decision-tree based classification technique to assess the degree of water contamination, based on the statistics of Government of India on water contamination, for six years from 2012-13 to 2017-18 (*www.data.gov.in and indiawaters.gov.in*). This statistics portrays the quality of water, based on the presence of Iron, Arsenic, Fluoride, Nitrate and Salinity. The proposed classification primarily classifies the pattern of contamination of each State in the India during specified period, in four point scale. Secondly, it classifies the quality of water into either potable or non-potable. This classification finds place in water quality assessment at regional and national level.

***Keywords*—** Data Analytics, Decision Tree, Classification, Water Quality Assessment

I. INTRODUCTION

Water is a natural Resource and it is indispensable for human life. But it naturally contains harmful minerals. Such as Iron, Nitrate, Salinity Arsenic and Fluoride, which are affecting human health. In water, minerals contamination level is high; it causes several health problems such as. If Arsenic level is high, it will create skin cancer, lung, liver and bladder problems; High Fluoride contamination causes bone density problems; High quantity of Iron in water, which creates pancreatic damage, diabetes, fatigue, weight loss, and joint pain in human; High Salinity increases the blood pressure and the Nitrate generates the blue baby effects. The authors have six years of data sets of minerals water contamination in all states of India from (*www.data.gov.in and indiawaters.gov.in*). Data sets of total number of Minerals (Iron, Arsenic, Nitrate, Fluoride, and Salinity) found in all states in India. Water contamination of the minerals in all states are classified using the Decision Tree classification. The mechanism of which selects the minerals individually from data set Find the progress of minerals and determine the four scale point values and three counting values.

II. RELATED WORK

Decision Tree Classification:

Decision tree builds classification models in the form of a tree structure. It divides the dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. Decision tree induction [1] can be used for evaluating features in the validation stage [2]. Decision-tree-based scheme is tested and trained datasets [3]. Decision tree classification approach is a low cost communication strategy using a set of most useful inter-base links for the classification task [4]. This classification model is used to prevent the SQL injection attacks [5]. Decision tree Classification (DTC) approach was developed to automatically extract urban land based on spectral and geographic features from Landsat TM images [6].Decision tree is a widely used technique to discover patterns from consistent dataset [7]. It is to create a decision tree that corresponds to the sample data [8]. The decision tree classification achieves the advantages of visual expression directly. Therefore decision tree is widely used in data analysis [9]. The reliability of decision tree is improved via attribute reduction [10]. It provides better classification that is presented in [11].

The proposed method segregated the given water contaminated minerals in six year dataset using decision tree based classification techniques.

III. METHODOLOGY

Decision Tree based Change Detection Analysis Method

This paper, proposes Decision Tree based change detection analytics method in water contamination minerals with assessment tree classification. For this analysis, the state wise six year water contamination data set of India is used, which is for the years between 2012 - 2018. Data set (are taken from www.data.gov.in and indiawaters.gov.in) contain five minerals such as Iron, Arsenic, Fluoride, Nitrate and Salinity. Rearrange the minerals in year wise as a separate excel file which in turn efficient for classification. The decision tree classification works on the ordered data sets. Difference is calculated among the minerals value in previous year and the Next year values consecutively. The differentiated values are stored into another excel sheet for further classification. With the differentiated data sets values the positive difference indicated by2, the negative difference indicated by 1,and the no difference indicated by 0.These indicated values are separately stored in another excel sheet for auxiliary classification. In additional classification, provide four scale points they are Increase, Decrease, No change, and Irregular. Further, the Increase count, Decrease count, and No change count are also computed.

The algorithm to find four scale point classification it shown below

Input: Water contamination minerals data sets.

Output: Four scale point classification report.

Algorithm for water contamination minerals classification

Step1: Import the minerals data sets.

Step2: Decision tree classification is used to segregate the five minerals.

Step3: Find the variations of same mineral for the two consecutive years.

Step4: Enumerate the variation scores or values.

Step5: Estimate the severity of variation by using four scale points

Step6: Count the Increase, Decrease, No change value.

Step7: Generate the water contamination severity classification report.

IV RESULTS AND DISCUSSION

Flow chart for water contaminated Minerals shows in Fig. 1. Table 1 shows on the water contaminated minerals founded places and its place count value for all states in India consecutive six year data sets. Table 2 shows the minerals here ranged in year wise.

Table 3 shows the Difference values. Differences are calculated among the minerals values with previous year and next year. It provides three values they are Positive difference, Negative difference, and No change difference. For example State Assam year 2012-2014 it provide negative difference,year 2013-2015 provide a positive difference, Another example Andhaman provide a no change difference.

Table 4 show the enumerated value. Negative difference are enumerated by 2, Positive difference enumerated by 1. And No difference enumerated by 0.

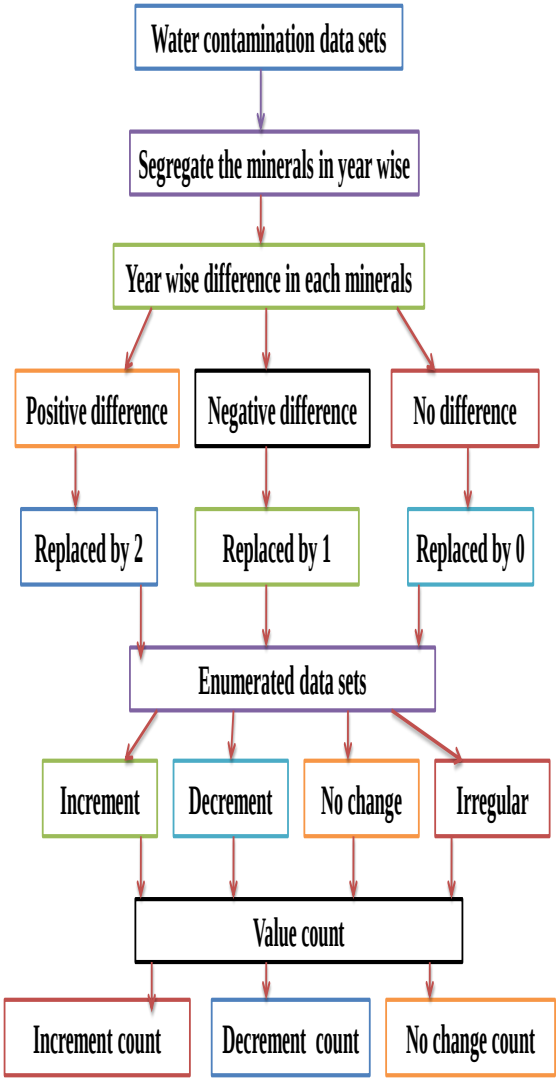

Fig. 1 Decision Tree Classification for Change Detection on Water Contamination Minerals Dataset

Table 1 Water contaminated Minerals

S. No	State Id	Arsenic	Fluoride	Iron	Salinity	Nitrate
1	S1	0	0	0	0	0
2	S2	2	3904	560	1922	182
3	S3	3	0	3	0	2
4	S4	1106	31	17174	2	0
5	S5	251	959	4308	6	7
6	S6	1	85	1214	24	12
7	S7	0	0	0	0	0
8	S8	0	1934	0	1068	3071
9	S9	0	403	8	65	7
10	S10	0	0	4	0	0
11	S11	0	5	7	0	1
12	S12	4	163	2607	2	11
13	S13	2	2254	1032	212	328
14	S14	0	114	889	22	144
15	S15	2	1369	346	195	342
16	S16	1	689	712	710	6515
17	S17	0	0	1	0	0
18	S18	0	0	4	0	0
19	S19	0	2	7	0	0
20	S20	0	0	0	0	0
21	S21	0	434	11744	300	0
22	S22	0	0	0	0	0
23	S23	802	357	256	15	127
24	S24	0	4542	0	2044	1909
25	S25	0	0	0	0	0
26	S26	1	876	220	1749	124
27	S27	0	3545	17	526	184
28	S28	0	0	716	0	0
29	S29	0	0	0	0	0
30	S30	1	3	4	0	0
31	S31	6677	393	18115	58	0

Table 2 for Arsenic

S. No	State Id	Arsenic 12_13	Arsenic13_14	Arsenic14_15	Arsenic15_16	Arsenic16_17	Arsenic17_18
1	S1	0	0	0	0	0	0
2	S2	2	0	0	0	2	0
3	S3	3	0	7	0	0	341
4	S4	1106	97	275	161	4268	3706
5	S5	251	628	247	282	844	1065
6	S6	1	5	0	0	30	0
7	S7	0	0	0	0	0	0
8	S8	0	1	0	0	0	0
9	S9	0	0	0	5	24	45
10	S10	0	6	0	0	0	0
11	S11	0	3	3	1	32	7
12	S12	4	3	134	12	4	126
13	S13	2	23	26	31	70	20
14	S14	0	0	7	0	2	3
15	S15	2	0	0	1	0	0
16	S16	1	2	22	0	4	1
17	S17	0	0	1	0	2	0
18	S18	0	0	0	1	45	1
19	S19	0	0	1	0	0	0
20	S20	0	0	0	0	0	0
21	S21	0	0	0	3	0	2
22	S22	0	0	0	0	0	0
23	S23	82	137	186	208	802	475
24	S24	0	0	0	0	0	3
25	S25	0	0	0	0	0	0
26	S26	1	0	1	0	0	0
27	S27	0	0	0	0	0	0
28	S28	0	0	0	0	0	0
29	S29	0	31	283	40	612	262
30	S30	1	0	1	0	1	0
31	S31	6677	2229	3644	6787	10552	7479

Table 3 Variation calculated table

S. no	State Id	Diff 12_14	Diff 13_15	Diff 14_16	Diff 15_17	Diff 16_18
1	S1	0	0	0	0	0
2	S2	2	0	0	-2	2
3	S3	3	-7	7	0	-341
4	S4	1009	-178	114	-4107	562
5	S5	-377	381	-35	-562	-221
6	S6	-4	5	0	-30	30
7	S7	0	0	0	0	0
8	S8	-1	1	0	0	0
9	S9	0	0	-5	-19	-21
10	S10	-6	6	0	0	0
11	S11	-3	0	2	-31	25
12	S12	1	-131	122	8	-122
13	S13	-21	-3	-5	-39	50
14	S14	0	-7	7	-2	-1
15	S15	2	0	-1	1	0
16	S16	1	-20	22	-4	3
17	S17	0	-1	1	-2	2
18	S18	0	0	-1	-44	44
19	S19	0	-1	1	0	0
20	S20	0	0	0	0	0
21	S21	0	0	-3	3	-2
22	S22	0	0	0	0	0
23	S23	665	-49	-22	-594	327
24	S24	0	0	0	0	-3
25	S25	0	0	0	0	0
26	S26	1	-1	1	0	0
27	S27	0	0	0	0	0
28	S28	0	0	0	0	0
29	S29	-31	-252	243	-572	350
30	S30	-1	-1	1	-1	1
31	S31	4448	-1415	-3143	-3765	3073

Table 4 Enumerated table

1	S1	0	0	0	0	0
2	S2	1	0	0	2	1
3	S3	1	2	1	0	2
4	S4	1	2	1	2	1
5	S5	2	1	2	2	2
6	S6	2	1	0	2	1
7	S7	0	0	0	0	0
8	S8	2	1	0	0	0
9	S9	0	0	2	2	2
10	S10	2	1	0	0	0
11	S11	2	0	1	2	1
12	S12	1	2	1	1	2
13	S13	2	2	2	2	1
14	S14	0	2	1	2	2
15	S15	1	0	2	1	0
16	S16	2	2	1	2	1
17	S17	0	2	1	2	1
18	S18	0	0	2	2	1
19	S19	0	2	1	0	0
20	S20	0	0	0	0	0
21	S21	0	0	2	1	2
22	S22	0	0	0	0	0
23	S23	1	2	2	2	1
24	S24	0	0	0	0	2
25	S25	0	0	0	0	0
26	S26	1	2	1	0	0
27	S27	0	0	0	0	0
28	S28	0	0	0	0	0
29	S29	2	2	1	2	1
30	S30	1	2	1	2	1
31	S31	1	2	2	2	1

Table 5 Minerals counting Table

Sno	State	12| 14	13 | 15	14 | 16	15 | 17	16 | 18	Status	Inc count	Dec count	No count
1	S1	0	0	0	0	0	No change	_	_	5
2	S2	1	0	0	2	1	Irregular	1	2	2
3	S3	1	2	1	0	2	Irregular	2	2	1
4	S4	1	2	1	2	1	Irregular	2	3	0
5	S5	2	1	2	2	2	Irregular	4	1	_
6	S6	2	1	0	2	1	Irregular	2	2	1
7	S7	0	0	0	0	0	No change	_	_	5
8	S8	2	1	0	0	0	Irregular	1	1	3
9	S9	0	0	2	2	2	Irregular	3	_	2
10	S10	2	1	0	0	0	Irregular	1	1	3
11	S11	2	0	1	2	1	Irregular	2	2	1
12	S12	1	2	1	1	2	Irregular	2	3	0
13	S13	2	2	2	2	1	Irregular	4	1	_
14	S14	0	2	1	2	2	Irregular	3	1	1
15	S15	1	0	2	1	0	Irregular	1	2	2
16	S16	2	2	1	2	1	Irregular	3	2	_
17	S17	0	2	1	2	1	Irregular	2	2	1
18	S18	0	0	2	2	1	Irregular	2	1	2
19	S19	0	2	1	0	0	Irregular	1	1	3
20	S20	0	0	0	0	0	No change	_	_	5
21	S21	0	0	2	1	2	Irregular	2	1	3
22	S22	0	0	0	0	0	Irregular	_	_	5
23	S23	1	2	2	2	1	Irregular	3	2	_
24	S24	0	0	0	0	2	Irregular	1	_	3
25	S25	0	0	0	0	0	No change	_	_	5
26	S26	1	2	1	0	0	Irregular	1	2	2
27	S27	0	0	0	0	0	No change	_	_	5
28	S28	0	0	0	0	0	No change	_	_	5
29	S29	2	2	1	2	1	Irregular	3	2	_
30	S30	1	2	1	2	1	Irregular	2	3	_
31	S31	1	2	2	2	1	Irregular	3	2	_

Table 5 result the status and pattern Hidden values such as Inc count, Dec count, and No count values .If the count value of all six years are 0, the status is updated as No change. The resultant values includes the combination of 0, 1 ,2 (or) 1,2,the status is updated as Irregular. The rest of the attributes, Such as Increase count, Decrease count, and No change count .shows the count values individually.

Note :

Table 6

State Id	State Name
S1	Andaman & Nicobar
S2	Andhra Pradesh
S3	Arunachal Pradesh
S4	Assam
S5	Bihar
S6	Chhattisgarh
S7	Goa
S8	Gujarat
S9	Haryana
S10	Himachal Pradesh
S11	Jammu & Kashmir
S12	Jharkhand
S13	Karnataka
S14	Kerala
S15	Madhya Pradesh
S16	Maharashtra
S17	Manipur
S18	Meghalaya
S19	Mizoram
S20	Nagaland
S21	Odisha
S22	Puducherry
State Id	**State Name**
S23	Punjab

S24	Rajasthan
S25	Sikkim
S26	Tamil Nadu
S27	Telangana
S28	Tripura
S29	Uttar Pradesh
S30	Uttarakhand
S31	West Bengal

Table-6 List of States

IV. CONCLUSION

In this paper, the Change Detection and Analysis on the water contamination minerals statistics is reported. The proposed method provides a classification pattern for Water Quality Assessment. This analytics would facilitate Government, Non-Government, Voluntary Organization and Industries on strategic planning and Polices Formulation.

References

[1] Laviniu Aurelian Bădulescu, *"Data mining classification experiments with decision trees over the forest covertype database"*, 21st International Conference on System Theory, Control and Computing (ICSTCC), pp.236, 2017

[2] M Niyas, K Sunitha, *"Decision tree learning based feature evaluation and selection for image classification"*, International Conference on Machine Learning and Cybernetics (ICMLC), pp.569,2017.

[3] A. Swetapadma, A. Jamehbozorg, Chengxi Liua, *"Identification and classification of fault during power swing using decision tree approach"*, IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Vol.1, pp.1, 10.1109/SPICES.2017.8091347, 2017.

[4] Tahar Mehenni, *"Integration of useful links in distributed databases using decision tree classification"*, International Conference on Information Systems and Economic Intelligence (SIIE) ,pp. 5,2015.

[5] *B. Hanmanthu; B. Raghu Ram; P. Niranjan*,"SQL Injection Attack prevention baed on decision tree classification", *IEEE 9th International Conference on Intelligent Systems and Control (ISCO),pp.1-5 2015.*

[6] L. Hua, W. Man, and Q. Wang, X.Zhao, "*A New Decision Tree Classification Approach for Extracting Urban Land from Landsat TM in a Coastal City, China"*, Fourth International Symposium on Information Science and Engineering, pp.282, 2012.

[7] M. Azad, M. Moshkov, *"Classification and optimization of decision trees for inconsistent decision tables represented as MVD tables"*, Federated Conference on Computer Science and Information Systems (FedCSIS), pp.31, 2015.

[8] D.A.Dlaeen, A. Alashqur, *"Using decision tree classification to assist in the prediction of Alzheimer's disease"*, 6th International Conference on Computer Science and Information Technology (CSIT), pp.122, 2014.

[9] P.Ye, "*The decision tree classification and its application research in personnel management"*, Proceedings of International Conference on Electronics and Optoelectronics, Vol.1, pp.372, 2011.

[10] W.Hui, W.Jing, Z. Tao, *"Analysis of decision tree classification algorithm based on attribute reduction and application in criminal behavior"*, 3rd International Conference on Computer Research and Development, Vol.1, pp.27,2011.

[11] P.L.Tu, J.Y. Chung, "*A new decision-tree classification algorithm for machine learning"*, Tools with Artificial Intelligence, TAI 92,Proceedings.,Fourth International Conference ,pp.370,1992.

Authors Profile

Dr. P. Shanmugavadivu pursed Bachelor of Science from K.N. Govt. Arts College, Thanjavur, 1987. She pursed Master of Computer Science and Applications from Regional Engineering College, Trichy, 1990 and Ph.D from The Gandhigram Rural Institute, India, in year 2008. She pursed Master of Business Administration from IGNOU, 2016. She is currently working as Professor in Department of Computer Science and Applications, Gandhigram Rural Institute, Dindigul, Tamil Nadu, India. She has published more than 22 research papers in reputed international journals and 31 research papers in edited volumes. Her main research work focuses on Medical Image Analysis, Image Restoration, Image Enhancement, Image Segmentation, Content – Based Image Retrieval. She has 24 years of teaching experience and 15 years of Research Experience.

P.Kavitha pursed Bachelor of Science from Arulmigu PalaniAndavar college of arts and Culture,Palani,2014. She pursed Master of computer Application from The Gandhigram Rural Institute,2017.She is currently pursuing M.Phil.in Department of Computer Science and Applications,The Gandhigram Rural Institute, India in 2018.Her research work focuses on data Analysis.

Mr. S. Dhamodharan pursed Bachelor of Science from Government Arts College, Melur, India in 2010 and Master of Computer Applications from Thanthai Periyar Government Institute of Technology, Vellore, India in year 2013. He is currently pursuing Ph.D and working as Project Fellow in Department of Computer Science and Applications, The Gandhigram Rural Institute (Deemed to be University), India since 2017. His main research work focuses on Digital Image Proessing, Artificial Intellengence, Machine Learning and Medical Imaging, Data Mining. He has one years of Research Experience.

