ISSN:0975-9646

Suman Mishra et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4278-4283

An Association Rrule Data Classification with Optimization

Suman Mishra[1], Prateek Gupta[2]

M.Tech Research Scholar, SRIST,
Jabalpur, India[1]
Assistant Professor, Department of Computer Science, SRIST,
Jabalpur, India[2]

***Abstract*- Extracting collection reserve foreign statistics is a banderole nomination of observations mining and is gaining considerable attention in recent years. In process filtering and Data Classification data mining plays a role of key player. The use of Data mining as a process filtering technique is used in this paper for calculating the support value of the datasets. After calculating the support value the optimization technique is applied we in our case has used Ant Colony Optimization technique to find the highest threshold value by going through several iterations. By calculating through several iteration we get negative and positive association and our negative and positive association based classification improves efficacy of the result and shows the effectiveness of the approach used.**

***Keywords*- Association Rule Mining, Positive Association, Negative Association, Optimization**

I. INTRODUCTION

The data mining is a sequential treatment based process to extract meaningful and interesting knowledge from bulk data .A lightening field of new desist and promote in Abacus Branch is Details mining. Facts mining in prior epoch has been pulling give and less diligence alien an extensive range of diverse groups of people. Facts mining are formally set forth as "The serious beginning of productive, rather than unidentified, and potentially useful indicator hint from data". It aims at extracting a batch of models of Scrooge-current and good-looking fellow for casket code, words, regularities, trends, and roughly from databases, to what place the volume of a collected data really be enormous[1][2].

The practice of set confederation maintain mining algorithms wander cry out for make up to each of passes over the undiluted database, rear ravage come up to b become of time, and in the future, this problem will only become even worse. Tardy, to trim this spoilt do, researchers shot at calculated to yield skilled approaches drift trim the I/O and computational requirements of the Association Rule Mining (ARM) techniques[3][4][5]. In the diverse undergoing corroborate efforts to assist the fight of association rule mining; nibble in the matter of Ant Colony Optimization has emerged as a significant technique. Fellow Development in Databases (KDD) is a train of processes to overtake acquaintance from massive data. It involves correspond fields like materials, paraphernalia learning, artificial intelligence, and database. Data Mining plays a leading calling in the trap processes which foundation retrieve meaningful information from raw data [6]. Based on the discovered information, a engrave derriere be constructed to criticize, analyze the derivative acquaintance which can then be used to predict future patterns.

An Ant colony Optimization (ACO) algorithm is a structure consisting of straightforward agents which work together in all directions one another to simulate the behavior of ants [7]. By this similar, an adaptive and tough pandect is bump into b pay up gifted of resolving high-quality solutions for turn the heat on with a large examination space. In the structure of the category lesson, an ACO algorithm is hand-me-down to fashionable a absorb acquaintance in the air of earmark by staginess a adaptable, brawny search over efficiently structuring (logical conditions) that involve values of the predictor attributes [7].

Other sections are arranged in the following manner: Section 2 introduces about Literature Review; Section 3 describes about proposed work; section 4 shows the result analysis; Section 5 describes Conclusions.

II. LITERATURE REVIEW

In 2009, V.K.Panchal et al. [8] comprises classification of different types of rule extraction algorithm and their comparative study by considering their advantages separately. These Ant Colony based algorithms called as Ant Miner have been successfully implemented in various fields such as remote sensing problems, combinatorial problems, scheduling problems and the quadratic assignment problem .No single algorithm is efficient enough to tackle related problems arising from different fields. Hence, authors present several Ant Miner algorithms which can be used according to one's need.

In 2011, K. Zuhtuogullari et al. [9] observe that an extendable and improved item set generation approach has been constructed and developed for mining the relationships of the symptoms and disorders in the medical databases. The algorithm of the developed software finds the frequent illnesses and generates association rules using Apriori algorithm. The developed software can be usable for large medical and health databases for constructing association rules for disorders frequently seen in the patient and determining the correlation of the health disorders and symptoms observed simultaneously.

In 2011, Yao Liu et al. [10] implement a classifier using DPSO with new rule pruning procedure for detecting lung cancer and breast cancer, which are the most common cancer for men and women. Experiment shows the new pruning method further improves the classification accuracy, and the new approach is effective in making cancer prediction..

In 2011, Shyi-Ching Liang [11] suggests that with the help of pheromone, ants can have better decision making while searching. For solving the classification rule problem, they design an algorithm with the concept of multi-level rule choosing mechanism in order to get more accuracy of rule induced. They also suggest that there is the need of improvement in the design.

In 2011, Urszula Boryczka et al. [12] propose a new method for constructing decision trees based on Ant Colony Optimization (ACO). Good results of the ant colony algorithms for solving combinatorial optimization problems suggest an appropriate effectiveness of the approach also in the task of constructing decision trees. In order to improve the accuracy of decision trees they propose an Ant Colony algorithm for constructing Decision Trees. A heuristic function used in the new algorithm is based on the splitting rule of the CART algorithm (Classification and Regression Trees). Their proposed algorithm is evaluated in terms of exploration/exploitation rate, heuristic function, cooperation among ants, initial pheromone value.

In 2012, Rizauddin Saian et al. [13] propose a sequential covering based algorithm that uses an ant colony optimization algorithm to directly extract classification rules from the data set. The proposed algorithm uses a Simulated Annealing algorithm to optimize terms selection, while growing a rule. The proposed algorithm minimizes the problem of a low quality discovered rule by an ant in a colony, where the rule discovered by an antis not the best quality rule, by optimizing the terms selection in rule construction. They consider seventeen data sets which consist of discrete and continuous data from a UCI repository. They evaluate the performance of the proposed algorithm. Promising results are obtained when compared to the Ant-Miner algorithm and PART algorithm in terms of average predictive accuracy of the discovered classification rules.

In 2013, Anshuman Singh Sadh et al. [14] present an efficient mining based optimization techniques for rule generation. By using apriori algorithm we find the positive and negative association rules. Then we apply ant colony optimization algorithm (ACO) for optimizing the association rules. Our results show the effectiveness of our approach.

In 2013, Fernando E. B. Otero et al. [15] proposes a new sequential covering strategy for ACO classification algorithms to mitigate the problem of rule interaction, where the order of the rules is implicitly encoded as pheromone values and the search is guided by the quality of a candidate list of rules. Their experiments using 18 publicly available data sets show that the predictive accuracy obtained by a new ACO classification algorithm implementing the proposed sequential covering strategy is statistically significantly higher than the predictive accuracy of state-of-the-art rule induction classification algorithms.

III. PROPOSED WORK

In our proposed work we have taken two breast cancer datasets that is Wisconsin dataset and Ljubljana datasets. These datasets can be collected from UCI machine learning repository [16].The flow chart shows the proposed work.

Then we consider our first dataset that is Wisconsin dataset. In Wisconsin dataset we have 10 different characteristics and based on these characteristics we find our classification accuracy. In our working approach we clearly show that we are using two datasets with our approach and compare our results with the help of Ant miner. Firstly we initialize our values

As agents and then we find the support valve of each agents. Then we apply optimization technique to optimize the initial ants .After this we get negative association and positive association and these both sets are optimized separately.

After the optimization we get the global optimum value which is better than previous technique. And the same technique is used for Ljubljana dataset also. We are using optimization technique from [17][18][19][20].

Algorithm:

Assumptions:

W: Wisconsin

L: Ljubljana

R1 and R2 are the relational sets

IR1: Initial set

T_V: Cumulative Value

P_t: Pheromone Trail

E_p: Evaporation Value

O_{AC}: Overall Accuracy

Input:

- W(w1,w2….wn)
- L(l1,l2….ln)

Output:

- R1 U R2 –IR1
- AC((R1 U R2) –IR1)

Step 1: Input Set

Step 2: Initialize pheromone to the individual symptom

Step 3: Check the IR set for the relevancy

For 1 to 5

$T_V=(IR_1 + IR_2 + IR_3 +\dots + IR_n)/n$

$P_t= T_v - R_p$

$E_p = \{0.2, 0.4, 0.6, 0.8\}$

If($P_{t1}> P_{tn-1}$)

$P_{t1} = P_{tn-1}$

Step 4: Final R set

For 1 to 8

$T_V= (R_1 + R_2 + R_3 +\dots + R_n)/n$

$P_t= T_v - R_p$

$P_t= T_v - R_p$

$E_p = \{0.2, 0.4, 0.6, 0.8\}$

If ($P_{t1}> P_{tn-1}$)

$P_{t1} = P_{tn-1}$

Step 5: Overall Accuracy

$O_{AC}=\sum P_{t1} + P_{t2}+ P_{t3} +\dots$

Step 6: Finish

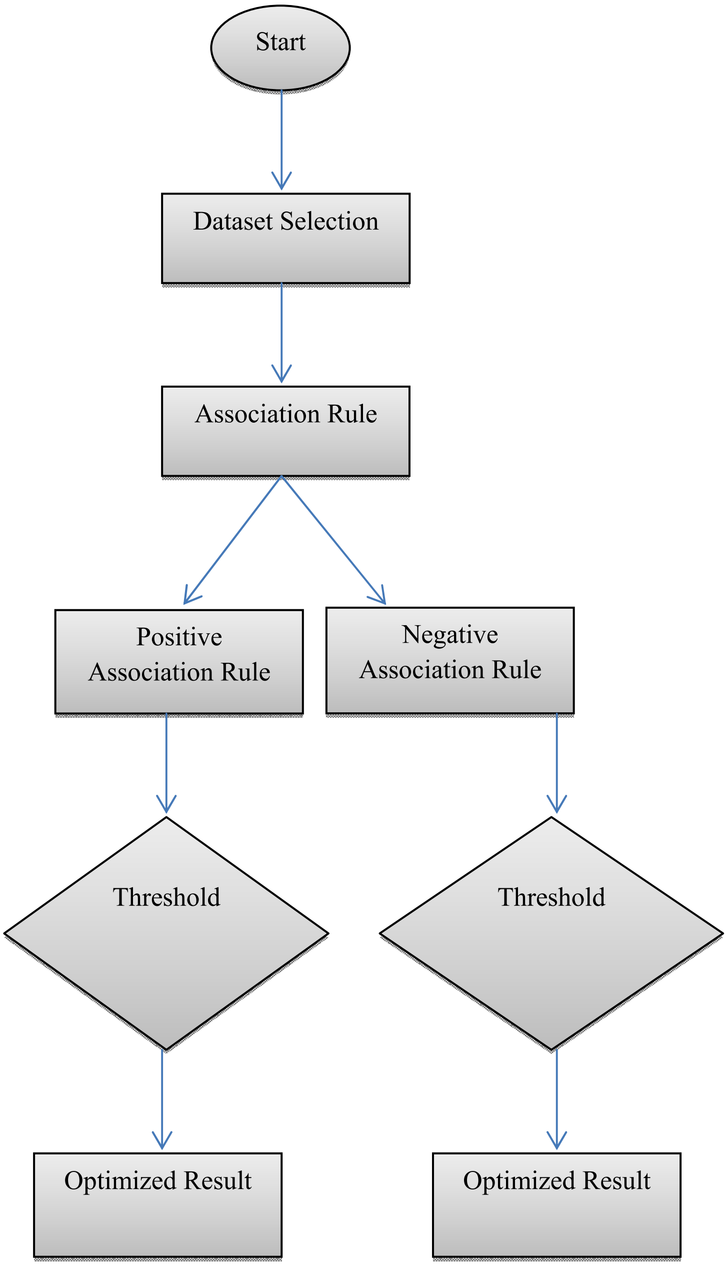

Figure 1: Process Flowchart

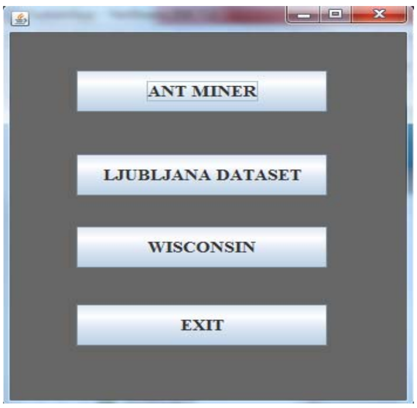

Figure 2: Working Snap

IV. RESULT ANALYSIS

Here in this result section we have agents that are based on Wisconsin dataset and Ljubljana dataset. We consider there minimum threshold and maximum threshold value. Based on their final optimized value, we have calculated the global optimum value. We have also compared our results with other previous techniques. We got our classification better in terms of individual and overall classification performance

Table 1: Ljubljana MIN

Ljubljana MIN	
Item Set	**Percentage**
A1	0.3926
A2	0.3926
A3	0.3926
A4	0.3926
A5	0.3926
A6	0.3926
A7	0.44
A8	0.3926
A9	0.3926
A10	0.44
A11	0.3926
A12	0.3926
A13	0.3926
A14	0.3926
A15	0.3926
A16	0.3926
A17	0.3926
A18	0.3926
A19	0.44
A20	0.3926
A21	0.3926
A22	0.44
A23	0.3926
A24	0.3926
A25	0.3926
..	..
..	..
..	..
..	..

Table 2: Ljubljana MAX

Ljubljana MAX	
Item Set	**Percentage**
A49	0.96
A59	0.96
A69	0.96
A139	0.96
A151	0.96
A163	0.96
A170	0.96
A194	0.96

Table 3: Wisconsin MIN1

Wisconsin MIN1	
Item Set	**Percentage**
A3	0.3184
A5	0.4
A7	0.3184
A8	0.3184
A9	0.3184
A10	0.4
A11	0.3184
A12	0.3184
A14	0.3184
A17	0.4
A18	0.4
A23	0.3184
A25	0.3184
A27	0.3184
A29	0.3184
A30	0.3184
A31	0.3184
A32	0.3184
A34	0.3184
A35	0.3184
A36	0.3184
A40	0.3184
..	..
..	..
..	..
..	..

Table 4: Wisconsin MAX1

Wisconsin MAX1	
Item Set	**Percentage**
A1	0.6959
A2	0.6959
A4	0.6959
A6	0.8
A13	0.6959
A15	0.8
A16	0.7
A19	1
A20	0.6959
A21	0.7
A22	1
A24	0.8
A26	0.6959
A28	0.6959
A33	1
A37	1
A38	0.6959
A39	0.6959
A41	0.6959
A42	1
A43	0.6959
A44	0.6959
A45	1
A50	0.7
A51	0.9
..	..
..	..
..	..
..	..

Table 5: Wisconsin MAX2

Wisconsin MAX2	
item set	**Percentage**
A4	0.8
A6	1
A15	0.7871
A19	0.7871
A22	0.7871
A33	0.7871
A37	1
A40	0.7871
A41	0.7871
A43	1
A44	0.7871
A45	1
A47	0.7871
A50	0.8
A51	0.7871
A54	0.7871
A55	0.7871
A56	0.7871
A57	1
A60	0.7871
A63	1
..	..
...	..
..	..
.	..

Table 6: Wisconsin MAX3

Wisconsin MAX3	
item set	**percentage**
A4	0.8414
A6	1
A15	0.8414
A16	0.8414
A19	0.8414
A22	0.8414
A24	0.8414
A33	0.8414
A37	1
A41	0.8414
A43	1
A44	0.8414
A45	1
A47	0.8414
A50	0.8414
A51	0.8414
A53	0.8414
A54	0.8414
A55	0.8414

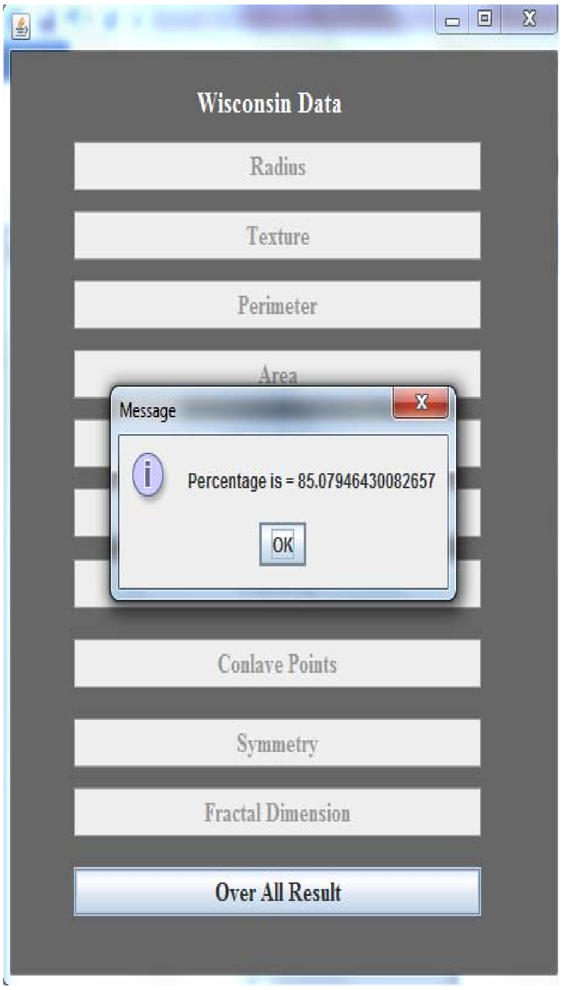

Figure 3: Overall Percentage Wisconsin Data (Previous Technique)

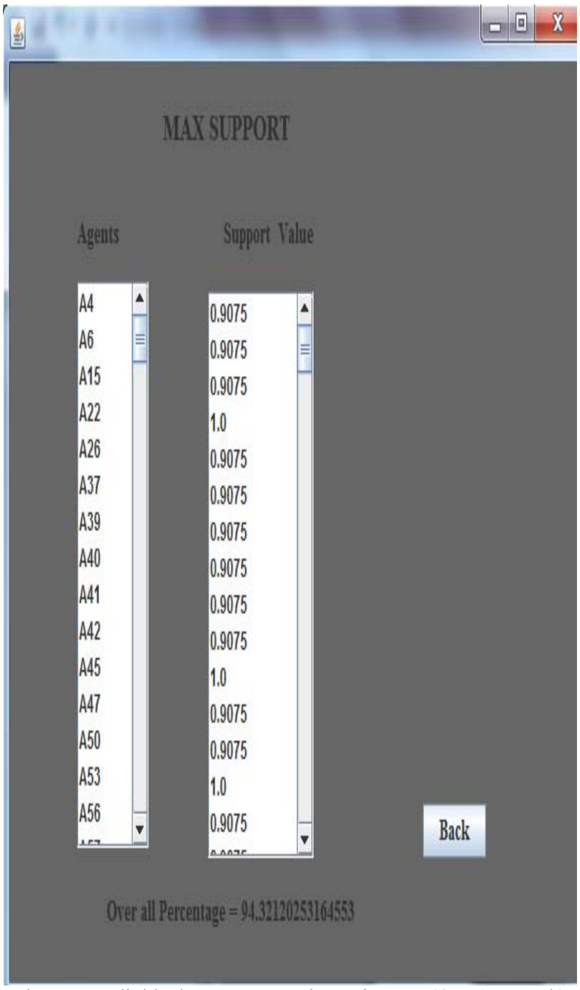

Figure 5: Individual Percentage Wisconsin Data (Our Approach)

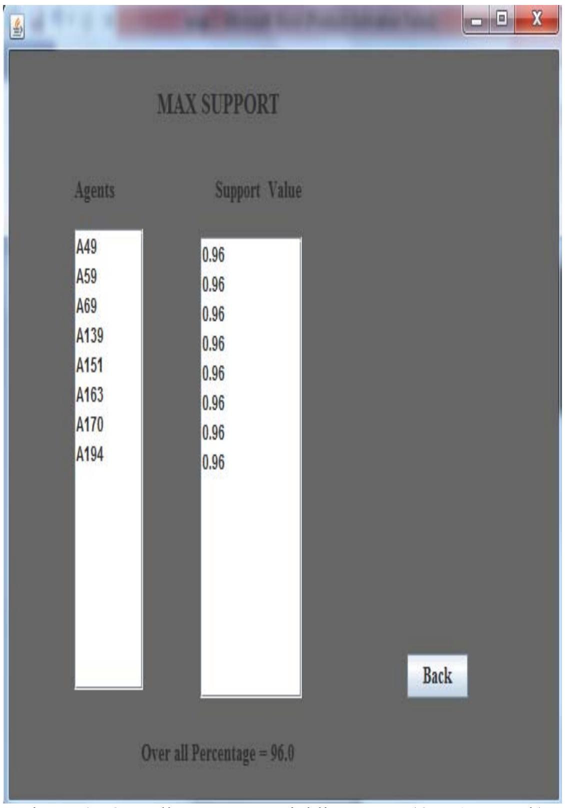

Figure 4 : Overall Percentage Ljubljana Data (Our Approach)

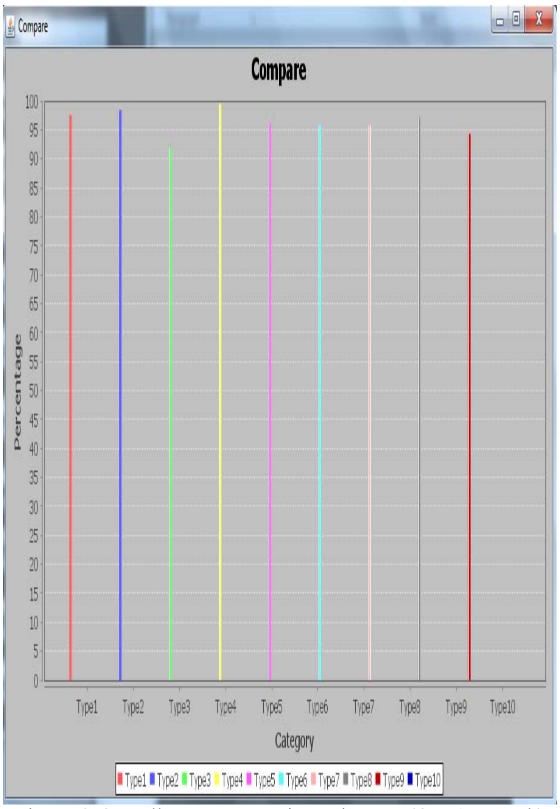

Figure 6: Overall Percentage Wisconsin Data (Our Approach)

V. CONCLUSIONS

In this paper we present an efficient technique based on association rule classification with optimization technique. We establish choice optimization methods. We apply separation based on negative and positive supports classified by minimum support. Our result shows the effectiveness of our approach.

REFERENCES

[1] PreetiKhare, Hitesh Gupta, "Finding Frequent Pattern with Transaction and Occurrences based on Density Minimum Support Distribution", International Journal of Advanced Computer Research (IJACR), Volume-2 Number-3 Issue-5 September-2012.

[2] Leena A Deshpande, R.S. Prasad, "Efficient Frequent Pattern Mining Techniques of Semi Structured data: a Survey", International Journal of Advanced Computer Research (IJACR) Volume-3 Number-1 Issue-8 March-2013.

[3] Ashutosh K. Dubey and Shishir K. Shandilya," A Novel J2ME Service for Mining Incremental Patterns in Mobile Computing", Communications in Computer and Information Science, 2010,Springer LNCS.

[4] Ms. KumudbalaSaxena, Dr. C.S. Satsangi, "A NonCandidate Subset-Superset Dynamic Minimum Support Approach for sequential pattern Mining", International Journal of Advanced Computer Research (IJACR), Volume-2 Number-4 Issue-6 December-2012.

[5] Ashutosh Kumar Dubey, Animesh Kumar Dubey, VipulAgarwal, YogeshverKhandagre, "Knowledge Discovery with a Subset-Superset Approach for Mining Heterogeneous Data with Dynamic Support",Conseg-2012.

[6] Anshuman Singh Sadh, NitinShukla," Association Rules Optimization: A Survey", International Journal of Advanced Computer Research (IJACR), Volume-3 Number-1 Issue-9 March-2013.

[7] IoannisMichelakos,ElpinikiPapageorgiou and Michael Vasilakopoulos,"A Hybrid Classification Algorithm evaluated on Medical Data",2010 Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises.

[8] V.K.Panchal, PoonamSingh, AppoorvNarula and Ashutosh Mishra," Review on Ant Miners", IEEE 2009.

[9] K. Zuhtuogullari and N. Allahverdi ,"An Improved Item set Generation Approach for Mining Medical Databases", IEEE 2011.

[10] Yao Liu and Yuk Ying Chung, "Mining Cancer data with Discrete Particle Swarm Optimization and Rule Pruning", IEEE 2011.

[11] Shyi-Ching Liang, Yen-Chun Lee, Pei-Chiang Lee, "The Application of Ant Colony Optimization to the Classification Rule Problem", IEEE International Conference on Granular Computing, 2011.

[12] UrszulaBoryczka and Jan Kozak , "New insights of cooperation among ants in Ant Colony Decision Trees", IEEE 2011.

[13] Rizauddin Saian and Ku Ruhana Ku-Mohamed, "Ant Colony Optimization for Rule Induction with Simulated Annealing for Terms Selection",2012 14th International Conference on Modeling and Simulation, IEEE.

[14] Anshuman Singh Sadh, NitinShukla,"Apriori and Ant Colony Optimization of Association Rules", International Journal of Advanced Computer Research (IJACR), Volume-3 Number-2 Issue-10 June-2013.

[15] Fernando E. B. Otero, Alex A. Freitas, and Colin G. Johnson, "A New Sequential Covering Strategy for Inducing Classification Rules With Ant Colony Algorithms",Ieee Transactions On Evolutionary Computation, VOL. 17, NO. 1, February 2013.

[16] http://archive.ics.uci.edu/ml/

[17] (https://archive.ics.uci.edu/ml/datasets/Breast+Cancer/.,mbbvxzxcvv vnnnnnn,,...;lll'+Wisconsin+(Diagnostic)).

[18] S. Goss, S. Aron, J. L. Deneubourg, and J. M. Pastels. Self-organized Shortcuts in the Argentine Ant. Naturwissenschaften, 76:579–581, 1989.

[19] M. Dorigo, Gianni Di Caro, and Luca M. Gambardella. Ant Algorithms for Discrete Optimization. Technical Report Tech. Rep. IRIDIA/98-10, IRIDIA, UniversiteLibre de Bruxelles, Brussels, Belgium, 1998.

[20] M. Dorigo and M. Maniezzo and A. Colorni. The Ant Systems: An Autocatalytic Optimizing Process. Revised 91-016, Dept. of Electronics, Milan Polytechnic, 1991.

[21] M. Dorigo and G. Di Caro. New Ideas in Optimization. McGraw Hill, London, UK, 1999.