International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

An Adaptive Slide Window Security Method for Transaction Updation in Data Stream Mining

Jayendra Kumar, Anitha Raju

Abstract*: Data steam mining has gained large interest in current research domain. Where various information's are retrieved based on the content of the context, the accuracy of the input stream with respect to its privacy is a major challenge. Windowing technique is used an effective approach in providing security measure in data stream mining. The recent develop windowing approach operates using sliding window, where anonymity is focused by different processing rules. The linear search sliding window has a constraint of search overhead and loss of generality under distributed information. In this paper, a new adaptive window approach for privacy coding in data stream mining is proposed. This presented approach is developed with the concern of minimize the search overhead and accuracy in search mining performance using adaptive window monitoring.

Keywords: Slide window approach, adaptive window coding, data stream mining.

I. INTRODUCTION

Data stream mining (DSM) has evolved as a major area of research in recent past. With large number of information been shared over the transaction, the preservation and accuracy of data stream has become a major concern. In data streaming publishing data need to maintain the privacy and should operate for publishing useful information on maintaining personal privacy. Recently, security in DSM has gained much attention in academics, industries, and various data publishing domain [1-3]. Here, Privacy implies data containing a details of transaction for identification and transaction details. With the emergence of large data streaming, data can be repeated in limited and growing level to tolerate the attacks [4]. Multilingual online applications such as network analysis [5] provide frequent transmission transactional data streams for security provision. However, disclosure of the data stream affects the privacy in the network. The privacy issue for publishing static transactional data has gained large interest in recent years. However, the nature of streaming and the approaches of privacy threatening has constraint the transaction performance. In [6] transactions are given with time period known as landmarks and are considered until the current time of transaction. In the process of sliding window the input data is processed with window given for data mining called as transactions-sensitive windows and time sensitive windows.

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
Jayendra Kumar*, Research Scholar, CSE, Koneru lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist,.A.P.India.Pin:522502,
Dr. Anitha Raju, Associate Professor, CSE, Koneru lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist.,A.P.India Pin:522502,

These windowing are applicable to data, as transactions lead to the removal of the transactional data for higher volume of data. To develop a privacy measure in transaction data in [7] a Model is presented, which set most of the background scenario by reducing a k-factor for related data. Here the Anonymity mechanism has been incorporated by monitoring and cooperating to privacy coding in streaming data . In [8-10] Background knowledge was transferred to clients, which defines for a transaction. In [11] an attacker is assumed to have an unregulated number of intrusion, and presented a monitoring system, which observe the groups, and then has sensitive values for each group defined to monitor. [12-16] defines a anonymity approach where data is defined in easy and accessible format rather than a non-formal recording using a single and generalization hierarchy. [17] Introduces a controlling approach which define the conversion, which share transactions with the maximum knowledge of materials and can be shared with a public privacy. In addition, they define global operation to protect several possible attacks for privacy constraints [18]. In all respects, the opponent's background knowledge was confined as object of observation. However, an attacker will receive partial knowledge of sensitive items and therefore, the idea of uncertainty in the privacy model will not allow an attacker to know the strategy used in defining the confidence of any section of data.

In Data stream operation data are persistent and ambiguous, and are usually unrelated [19]. Publishing data with security measure related methods are outlined in [20-23] for security provisioning in DSM. To introduce a unify data stream using k-anotomy [24] approach an aided input is defined for security application. In [58] for time constraint data publication a cluster reuse, and an approach of data streams was introduced by clustering based on the speed limit of the process with reduction in information loss. The concerns of information loss result in false reputation and decrease the reliability of the system. With objective to develop a security approach in data stream mining, in this paper, a new adaptive sliding window coding for security provision minimizing information loss is proposed. To outline the proposed work, this paper is outlined in 6 sections, where section 2 brief the approach of security coding in data stream using sliding window approach. Section 3 presents the proposed approach of adaptive slide mode coding, section 4 presents the simulation result developed for the proposed work, and section 5 concludes the presented work.

II. SECURITY MEASURE IN DATA STREAM MINING

Towards developing a security approach in DSM Anonymity issue for privacy preservation is outlined in [1]. Here window uncertainty is observed and a adaptive window sliding is made to satisfy the security constraint. The adaptive sliding is monitored via information metric and for each of the transactional sliding window, a information loss is computed by varying different window slide. A confidence factor is developed using a suppression algorithm satisfying the uncertainty constraint. However, the security measure are developed based on the transaction made and the window data stream monitor. The issue observed with the presented approach is the uncertainty of user access to the window where the access of transaction information loss is defined by the number of transaction made for a monitoring parameter. In [1] Two dynamic algorithms with generalization and suppression to anonymize using continuously sliding window are presented. To make it satisfy ρ-uncertainty by structuring an affected sensitive rules trie, because the removal and addition of transactions may make the current sliding window fail to satisfy ρ-uncertainty.

A transactional data stream based on sliding window is presented, where the window size and step size are defined for a window length. An illustration of window operation is shown in figure 1.

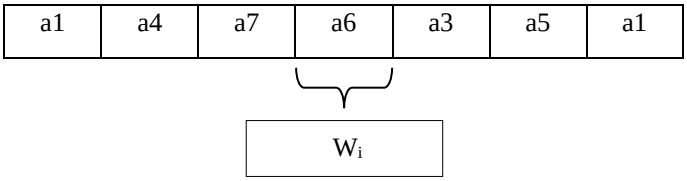

Fig 1: Window selection scheme in DSM

The transaction items *a*i are non-sensitive, and the parameter of window controlling is sensitive in privacy coding. Here, the transactional sliding window T_{SWd} contains transactions t_i…….t_{i+1}. When new transactions t_{i+1} arrive, the last 2 transactions are deleted from the transactional sliding window, and the current window is updated. This process is repeated for each 2 window information to derive privacy. An example of the slide window outlined in [1] is as given.

Assume user-1 knows that user-2 bought a_2 and b_1 from a store on one day and that the store publishes its shopping data stream for mining. It is also assumed that user-1 can monitor the data stream by sliding window, and knows the transaction of user-2 in T_{SWdi} e. Then, user-1 can infer that user-2 also bought, so user-2 privacy is compromised.

Assume that an attacker can monitor a data stream by sliding window model, and knows a subset of items σ of the victim's transaction t in a transactional sliding window. If an attacker can infer with high probability that t contains σ and it also contains a sensitive item from the transactional sliding window, the privacy is compromised. It means to mine the rule from the transactional sliding window, where σ is the antecedent and μ is the consequent. Such an rule is called a sensitive association rule. The confidence of a security system in a transactional sliding window T_{SWi} can be computed by sup( ), where sup( ) is the number of transactions in T_{SWi} which contains σ. To prevent these inferences, every transactional sliding window require that the coding appearing in a transactional sliding window is less than a threshold value ρ.

To avoid the generation of false rules, global suppression is used. When the sliding window is updated, the sliding window may not satisfy required length due to the removal and addition of transactions. Here, a continuous window sliding is used to handle it and make it satisfy ρ-uncertainty.

However, here the convergence condition is developed over sequential sliding window which as 2 constraint. Firstly, the approach has a 2 window search which has no consideration of past learnt observation, and secondly the windowing has a constraint search size limiting the accuracy of matching in stream mining. To overcome the state issue a adaptive window selection with past learning method is proposed.

III. ADAPTIVE WINDOW CODING

In the proposed approach, the adaptation of an window due to random data input is analyzed. Here the added transaction by a user W_{add} are randomly been updated, which are observed to give an additional security threat to the existing mining approach. In the updation of secondary transaction, here, the updation of the transaction weight is based on the current update and the past records. The W_{del} / W_{add} updation depends on the current update and the node performs a transaction evaluation operation to drive the past records. In the process of transaction updation , the updation is dynamically been updated or deletion based on the availability and current update. In the process of transaction evaluation or updation, the windowing area dynamically been switched and the window synchronized between the querying and testing value is selected in either a random selection mode or a full sequential mode. For an initial window updation of $w \in W$, where w is the allocated window and W is the total windows, wherein the window updation w^t is observed to updated with the updation or deletion of a transaction. The updation of the transaction windows is defined by,

$$w_{av1}^t = \sum_{n \in N} w_n^t - w_i^t \quad (3)$$

Where $\sum_{n \in N} w_n^t$ is the aggregated sum of all the transaction windows observed. Each Window process a transaction evaluation operation, based on the approach of repository search model , defined by;

$$R_s = \sum_{i=1}^{n} \left(\frac{T_i}{t}\right)^{\delta} \quad (4)$$

Where T_i is the transaction to observe. The repository threshold for the transaction evaluation is given by eqn. 4, where 'δ' is an governing factor to the probability of transaction variations.

In the process of transaction changes , the estimated repository of the updated transaction is compared with the derived match thresholds, and a transaction having repository higher than this threshold is taken as a varying transaction, else the system consider it as no transaction.

$$wupdt = \begin{cases} 1; & R_s \geq R_{limit} \\ 0; & otherwise \end{cases}$$

The total window updation is then given by,

$$w^t_{avl+} = \sum_{n \in N} w^t_n + w^t_{avl+} \quad (5)$$

Where, w^t_{avl+} is the observed transaction for the $wupdt$.

In the updation of this transaction , the search rate will be improved as the past transaction is observed. However, it is observed that, a random updation of this past transaction could lead to higher search overhead as the transaction is added with a higher volume of data. it is hence, optimized as a tradeoff for the update offered over the delay rate. This problem is solved as a maximum transaction rate updation problem. This approach optimizes the transaction updation based on the constraint of minimal delay and maximum records updation . A cost function in this case is defined by,

$$w^t_{alloc} = \min_{\{n\}}\{t\} \Rightarrow w^t_{avl+} \leq w^t_{avl+} \quad (6)$$

Here, from the observed transaction , the allocable transaction w^t_{alloc} is constraint with the delay minimization problem. It is computed for the past transaction to observe the delay occurrence, and for the w^t_{alloc} giving minimum delay $\min_{\{n\}}\{t\}$ is selected. This updation guarantees the offering higher security of data in update, with higher transaction updation . The proposed transaction updation approach is outlined in the given algorithm below,

Algorithm:

Sense the past transaction w^t_{avl+} *for the period t;*
Formulate a repository R;
Calculate the Delay Repository (D_{int})
if ($D_{int} < D_w$)
Generate a transaction selection window step μ*;*
if ($w^t_i = \mu$)
set the transaction window;
else
$w^t_i = w^t_{i+1}$
End

Wherein in the scenario of random transaction updation and additional transaction resulting in lower transaction delay which can be used for data exchange resulting in higher secure updation. However, the pooling of transaction updation in a random updation is higher resulting in delay. This is minimized by selecting a optimal window width in the proposed approach.

IV. SIMULATION RESULTS

The validation of the proposed work is performed by the analysis of information loss and delay metric over different data base entry, and the values of the window size *w*, step size *p* and privacy requirement ρ. The test system is tested over BMS-WebView-2 database [25]. The packet loss is computed as the number of updation been made over the total transactions for a time period. The system delay is measured to derive the computational delay taken in computing the process

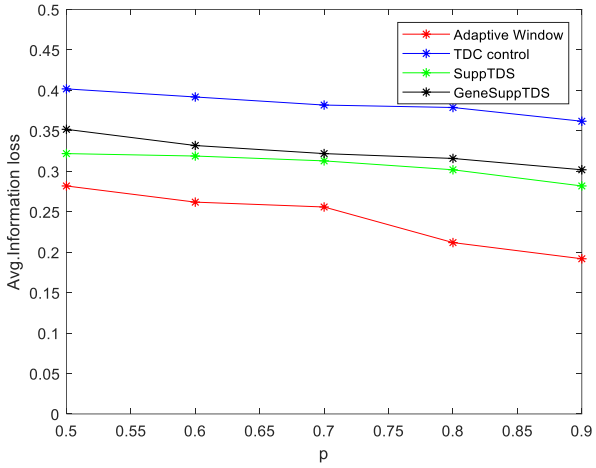

Fig 2: Average packet loss for different value of step size p

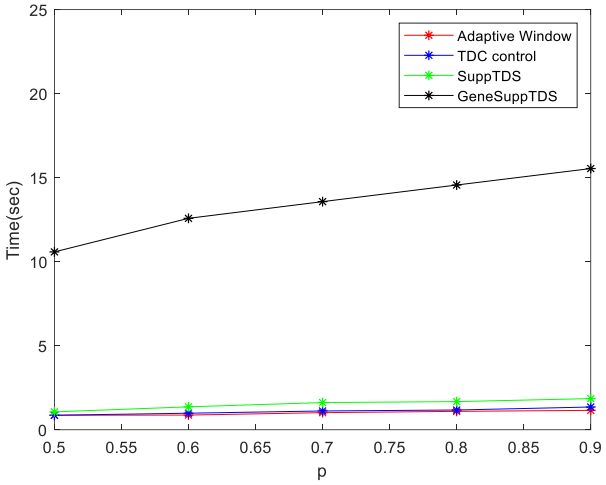

Fig 3: Delay factor for different value of step size p

Observation for the measured delay and packet loss parameter for variation in step updation size is presented in figure 2 and 3 respectively. The approach of adaptive windowing technique reduces the delay metric by suitably selecting a window size which best fit the transaction observation. The packet loss in this case is observed to minimal due to past repository monitoring. The past observation improves the updation or deletion of the transaction entry in the data base.

Table 1: Observation of information loss for varying p size

p	Adaptive Window	TDC control	SuppTDS	GeneSuppTDS
0.5	0.18	0.3	0.22	0.25
0.6	0.16	0.29	0.217	0.23
0.7	0.154	0.28	0.211	0.22
0.8	0.11	0.277	0.2	0.214
0.9	0.09	0.26	0.18	0.2

Table 2: Observation of delay (Sec) for varying p size

p	Adaptive Window	TDC control	SuppTDS	GeneSuppTDS
0.5	0.5	0.5	0.7	10.22
0.6	0.51	0.62	1.0	12.217
0.7	0.654	0.754	1.254	13.211
0.8	0.7311	0.811	1.311	14.20
0.9	0.79	0.99	1.49	15.18

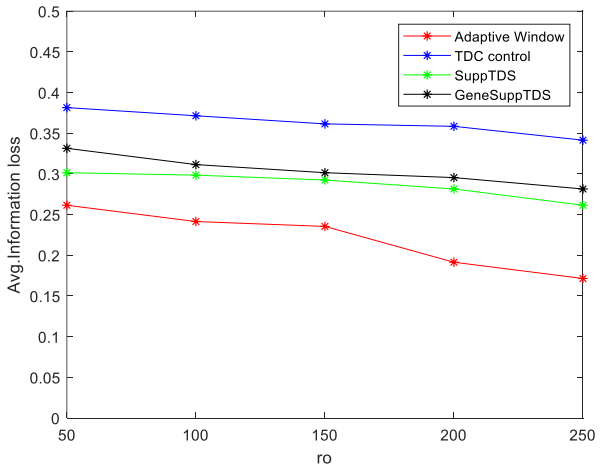

Fig 4: Average packet loss for different value of step size ρ

The observation of packet loss for varying ρ is presented in figure 4. It is observed that with increase in the value of ρ, the loss is minimized. This is observed to be minimal for the proposed adaptive window approach. The method control the windowing by the transaction entry and past observation which gives the selection approach a suitability in packet updation resulting in lower information loss.

Table 3: Observation of information loss for varying ρ size

ρ	Adaptive Window	TDC control	SuppTDS	GeneSuppTDS
50	0.23	0.34	0.3	0.34
100	0.2	0.33	0.31	0.33
150	0.18	0.325	0.308	0.32
200	0.15	0.328	0.27	0.28
250	0.13	0.31	0.25	0.25

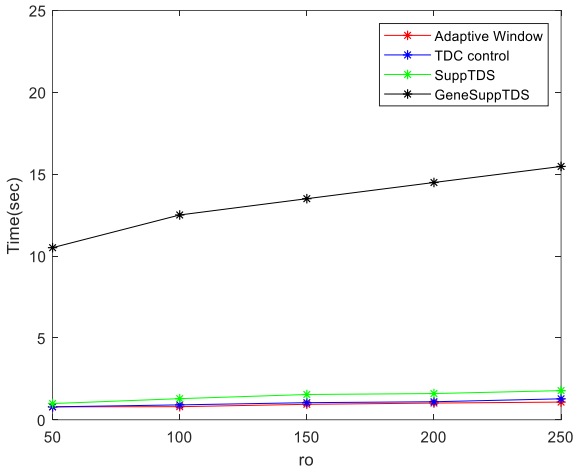

Fig 5: Delay factor for different value of step size p

Table 4: Observation of delay for varying ρ size

ρ	Adaptive Window	TDC control	SuppTDS	GeneSuppTDS
50	0.56	0.61	0.72	10.54
100	0.59	0.65	1.3	13.62
150	0.62	0.73	1.5	14.11
200	0.7	0.82	1.9	15.20
250	0.79	0.89	2.1	16.3

Varying the window length for different method the packet loss and delay metric is monitored as presented in figure 6 and 7 below.

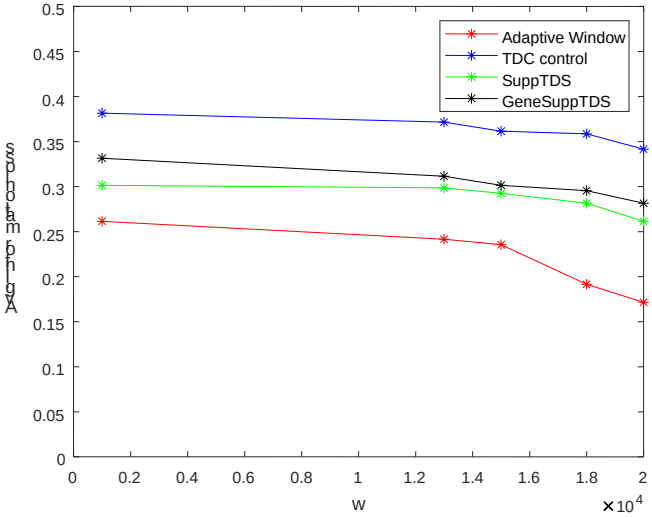

Fig 6: Average packet loss for different value of step size w

Retrieval Number: B3573129219/2019©BEIESP
DOI: 10.35940/ijeat.B3573.129219
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

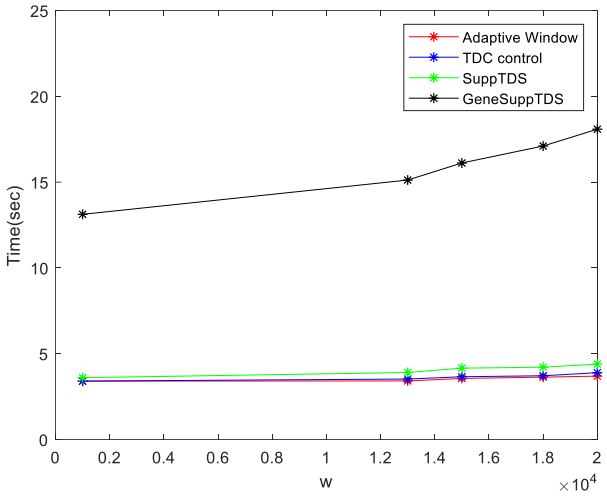

Fig 7: Delay factor for different value of step size w

Table 5: Observation of information loss for varying w size

w	Adaptive Window	TDC control	SuppTDS	GeneSuppTDS
1000	0.177	0.3	0.23	0.25
13000	0.163	0.289	0.24	0.223
15000	0.16	0.27	0.22	0.21
18000	0.1	0.273	0.21	0.22
20000	0.09	0.26	0.2	0.213

Table 6: Observation of delay for varying w size

w	Adaptive Window	TDC control	SuppTDS	GeneSuppTDS
1000	0.32	0.53	0.73	10.4
13000	0.41	0.66	1.2	12.7
15000	0.54	0.74	1.34	13.1
18000	0.61	0.81	1.41	14.0
20000	0.72	0.9	1.5	15.8

V. CONCLUSION

The proposed approach developed a new dynamic weight updation process in data stream mining. The constraint of linear search window with 2 selected window result in data lost and delay is observed. To minimize the delay metric and information loss, a new coding approach based on adaptive window selection and past referral model is developed. The approach of adaptive window selection and past repository monitoring resulted in minimization search delay and improve the system performance. The contributed lesser delay parameter could contribute to higher classification performance due to lower processing time.

REFERENCES

1. Jinyan Wang , Chaoji Deng, and Xianxian Li, "Two Privacy-Preserving Approaches for Publishing Transactional Data Streams", Special Section On Recent Computational Methods in Knowledge Engineering and Intelligence Computation, IEEE Access, Vol. 6, pp- 23648- 23658, 2018.
2. S. Wang, L. Minku, and X. Yao, "A learning framework for online class imbalance learning," in Proc. IEEE Symp. Comput. Intell. Ensemble Learn., Apr. 2013, pp. 36–45.
3. S. Wang, L. L. Minku, and X. Yao, "Online class imbalance learning and its applications in fault search," Int. J. Comput. Intell. Appl., vol. 12, no. 4, pp. 1340001(19 pages), 2013.
4. J. Kivinen, A. Smola, and R. Williamson, "Online learning with kernels," IEEE Trans. Transaction Process., vol. 52, no. 8, pp. 2165–2176, Aug. 2004.
5. N. Japkowicz, "Concept-learning in the presence of between-class and within-class imbalances," in Proc. 14th Biennial Conf. Can. Soc. Comput. Stud. Intell.: Adv. Artif. Intell., 2001, pp. 67–77.
6. T. Jo and N. Japkowicz, "Class imbalances versus small disjuncts," SIGKDD Explor. Newsl., vol. 6, no. 1, pp. 40–49, Jun. 2004.
7. P. Mallapragada, R. Jin, and A. Jain, "Non-parametric mixture models for clustering," in Proc. Int. Conf. Struct., Syntactic, and Statistical Pattern Recog., 2010, vol. 6218, pp. 334–343.
8. K. Bache and M. Lichman. (2013). UCI machine learning repository [Online]. Past: http://archive.ics.uci.edu/ml
9. R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang, "Towards social user profiling: Unified and discriminative influence model for inferring home locations," in Proc. 18th ACM SIGKDD Int. Conf. Know. Discovery Data Mining, 2012, pp. 1023–1031.
10. H. He and E. Garcia, "Learning from imbalanced data," IEEE Trans. Know. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009.
11. S. Wang, L. Minku, and X. Yao, "Resampling-based ensemble methods for online class imbalance learning," IEEE Trans. Know. Data Eng, vol. 27, no. 5, pp. 1356–1368, May 2015.
12. I. Ozalp, M. E. Gursoy, M. E. Nergiz, and Y. Saygin, ``Privacy-preserving publishing of hierarchical data," *ACM Trans. Privacy Secur.*, vol. 19, no. 3, Sep. 2016, Art. no. 7.
13. Y. Xin, Z.-Q. Xie, and J. Yang, ``The privacy preserving method for dynamic trajectory releasing based on adaptive clustering," *Inf. Sci.*, vol. 378, pp. 131_143, Feb. 2017.
14. H. Zakerzadeh, C. C. Aggarwal, and K. Barker, ``Managing dimensionality in data privacy anonymization," *Knowl. Inf. Syst.*, vol. 49, no. 1, pp. 341_373, Oct. 2016.
15. R. Chen, N. Mohammed, B. C. Fung, B. C. Desai, and L. Xiong, ``Publishing set-valued data via differential privacy," in *Proc. VLDB*, Seattle, WA, USA, 2011, pp. 1087_1098.
16. J. Liu and K. Wang, ``Anonymizing transaction data by integrating suppression and generalization," in *Proc. PAKDD*, Hyderabad, India, 2010, pp. 171_180.
17. S. Ramírez-Gallego, B. Krawczyk, S. García, M.Wo¹niak, and F. Herrera, ``A survey on data preprocessing for data stream mining: Current status and future directions," *Neurocomputing*, vol. 239, pp. 39_57, May 2017.
18. S. K. Tanbeer, C. F. Ahmed, B.-S. Jeong, and Y.-K. Lee, ``Sliding windowbased frequent pattern mining over data streams," *Inf. Sci.*, vol. 179, no. 22, pp. 3843_3865, Nov. 2009.
19. Y. Zhu and D. Shasha, ``StatStream: Statistical monitoring of thousands of data streams in real time," in *Proc. VLDB*, Hong Kong, 2002, pp. 358_369.
20. Z. Farzanyar, M. Kangavari, and N. Cercone, ``Max-FISM: Mining (recently) maximal frequent itemsets over data streams using the sliding window model," *Comput. Math. Appl.*, vol. 64, no. 6, pp. 1706_1718, Sep. 2012.
21. J. Kim and B. Hwang, ``Real-time stream data mining based on CanTree and Gtree," *Inf. Sci.*, vols. 367_368, pp. 512_528, Nov. 2016.
22. F. Nori, M. Deypir, and M. H. Sadreddini, ``A sliding window based algorithm for frequent closed itemset mining over data streams," *J. Syst. Softw.*, vol. 86, no. 3, pp. 615_623, Mar. 2013.
23. H. Chen, L. Shu, J. Xia, and Q. Deng, ``Mining frequent patterns in a varying-size sliding window of online transactional data streams," *Inf. Sci.*, vol. 215, pp. 15_36, Dec. 2012.

24. H. Ryang and U. Yun, ``High utility pattern mining over data streams with sliding window technique," *Expert Syst. Appl.*, vol.57, pp. 214_231, Sep. 2016.
25. Z. Zheng, R. Kohavi, and L. Mason, ``Real world performance of association rule algorithms," in Proc. KDD, San Francisco, CA, USA, 2001, pp. 401-406.

AUTHORS PROFILE

Mr. Jayendra kumar, is currently Research Scholar at Computer Science and Engineering Department Koneru Lakshmaiah Education Foundation (Deemed to be University) Vaddeshwaram,Guntur Dist.,AP . He obtained M.Tech CSE from JNTU Hyderabad. His research interest are Data Mining, Internet of Things and Machine Learning . He is Life Member of Computer Society of India

Dr. Anitha Raju, received a Ph.D. degree in Image Processing and Pattern Recognition from Sri Padmavathi Mahila Visvavidhyalayam, Tirupathi.She is a Woman Scientist sponsor by DST from Govt. of India under the scheme of WOS-A. She is Assoc. Prof. Dept. of Computer Science and Engineering in Koneru Lakshmaiah Education Foundation. Her research interest is the Internet of Things, Image Processing and Machine Learning.