IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

A Rigorous Review on an Ensemble Based Data Stream Drift Classification Methods

R. C. Samant[1*], D. M. Thakore[2]

Dept. of Computer Engineering, Bharati Vidyapeeth (Deemed to be university) College of Engineering, Pune, India

Corresponding Author: ruchasamant25@gmail.com, Tel.: +91-9960450076

DOI: https://doi.org/10.26438/ijcse/v7i5.380385 | Available online at: www.ijcseonline.org

Accepted: 10/May/2019, Published: 31/May/2019

Abstract— incurrent era of big data, thedata analytics has become more challenging issue. Data get mined for finding facts as well as to predict impact of various activities which is used everywhere in the life. Mining processes like classification and clustering becomes more crucial in case of dynamic data streams. As the nature of data stream is temporal there is always a difference in the concepts which causes a concept drift. This concept drift affects the reliability of classifiers and clustering methods. Classification is important technique in data mining, which has been applied with various modifications to handle concept drift issue. Data stream classification is different from normal classification process as it has restriction of time, memory size and speed of processing along with accuracy. This article presents a review of remarkable recent ensemble based classifiers, which are designed to detect concept drift based on the different way of data stream processing.

Keywords—Ensemble Learning, Data Stream Mining, Concept Drift.

I. INTRODUCTION

Big data is a mixture of a large volume of organized, semi organized and unorganized data that can be mined to extract knowledge from it [1]. Many applications such as social media, sensor networks, market analysis, internet communications and online news generate big data, the data which grows unlimited like millions of records per day. Thus the rate of data generation is tremendous. Along with this the need of processing this voluminous amount of data is increased. It is quite essential to process this data online to find out meaningful information from the ocean of data.

Data generated by various internet applications form data streams. These may be simply attribute-value pair or it may be short text like tweets. Data stream mining is a way of deducting knowledge and acquiring useful and relevant information from ocean of data [2]. Classification of big evolving data is one of the data mining techniques which have shown its strength of prediction of ubiquitous and large volume of data [3].

Data stream always changes over time, there is no upper bound on data and it is temporal. Similarly it is produced at high speed. Considering all these parameters the data stream processing has many limitations. It must be processed in finite amount of time with limited space.

This paper will put focus on data stream classification techniques. After detailed study of various data stream classification methods, it is found that data stream classifiers are designed using ensemble based classifiers works more efficiently. Instead of rely on incremental and online learning methods which work on a single model, ensemble learning uses different method. It divides the data streams into small data parts so that it can be processed by classifiers separately. Each classifier output is combined to predict exact result or prediction. This strategy help to improve performance in handling large data, adaptation to concepts is fast, and reduced error rate, and speed up data processing by parallelizing it.

The paper is organized as follows. Section I contains the introduction of Data stream. Section II, focuses on issues related to data stream processing. In section III study of online ensemble methods is carried out. Section IV presents data stream classification using chunk based ensemble methods. As it is a study and review paper , Section IV concludes with study of different ensemble classifiers.

II. DATA STREAM CLASSIFICATION

In the constantly changing input stream, issues like single pass processing, large storage space, anomaly detection, novel class detection and concept drift detection are challenging. The input stream distribution might change over course of time and it could lead to conditions of concept drift

[4]. By the definition, the concept drift means the statistical properties of the target variable, which the model is trying to predict, change over time in unforeseen ways and because of this problem predictions become less accurate as time passes [5],[6]. Here problem of classifier design is complex as training data and test data are not same and it is independently distributed. Prediction of such information becomes less accurate as input changes. In recent years remarkable efforts are taken by researchers to develop adaptive learning algorithms to detect concept drift. The concept drift detection problem is very wide and has many categories under it. Though the issue is relatively new, still there are some adaptive learning algorithms that were proposed earlier.

There are various classification methods in literature such as decision trees, rule based methods, neural networks, naïve based which has shown tremendous performance in various applications. Mostly these techniques has been used on static data streams, where multiple passes over a data is possible, similarly data is also less compared with data streams. Stream Classifiers has to work on data in a single pass as it`s bound by time. The accuracy of data classifiers depends upon prediction of labels, and in stream data identification of proper label is very difficult. There are many algorithms has been designed by using different concepts like sliding window, drift detectors, ensemble methods, active learning methods.

III. DATA STREAM CLASSIFICATION USING ONLINE ENSEMBLE METHODS

The ensemble based classification method is capable to handle concept drift, since it can effectively adapt to the fast changes in the underlying data [6],[7],[8],[9]. In ensemble based classifier multiple classifiers are used to reveal new classes in data. Design of classifier can be either of type EC1, where all the classifiers use same algorithm or as EC2 where every classifier uses different methods [6]. As Ensemble methods integrated with drift detection algorithms has shown excellent performance in data stream processing, similarly in many cases flexibility is added with dynamic updates, such as deletion or addition of selected classifiers [10]. Ensemble based classifiers can handled main three problems of data stream: concept drift, efficiency and robustness. Figure 1 depicts the working of ensemble method where input data stream is divided into parts and each classifier works on it. Final decision is calculated either by using voting scheme or by using posterior probability or setting some threshold values. Broadly there are two ways to designed ensemble based classifiers: online ensembles and blocked based ensemble.

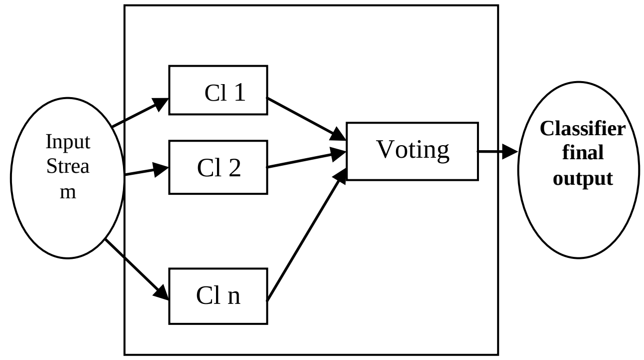

Fig. 1 Working of Ensemble based classifiers

The online ensemble learning method processes newly arriving instances without storing them, as an incremental learning method. Many researchers have proposed various online ensemble methods among them Online Bagging and Boosting method suggested by Oza and Russell [11],[12] is milestone in the history of data stream classification. In the online bagging method for training set Z, input is paired in (x, y). The numbers of iterations are fixed as well as machine learning method M. The machine learning technique may be decision tree, SVM, random forest or any other; it can change as per the requirement of data stream processing. For every iteration a bootstrap sample of size n is drawn from training set Z and a learning method M is applied on it. Each base classifier determines model f(x). The classification process gets repeated k times and then it adds the sample (x,y) to the training set and update all of the base classifiers, returning ensemble $\{f^1,\ldots.,f^B\}$. Classification of data stream is done by majority vote of all predictors. Online Boosting algorithm also works on samples repeatedly according to the Poisson distribution. It updates the subsequent base classifier if there is change in samples. Both classifiers works on data stream online, but not efficient in handling concept drift problem. Online version of bagging and boosting algorithm typically perform better in some data sets than batch processing method. These algorithms have low overhead and are quite suitable for practical applications [11].

Bagging and boosting are classical ensemble methods which improved the performance of a classifier but in context of revolving data stream there was a scope of improvement in these methods. A. Bifet, etc. proposed two new variants of Bagging: ADWIN Bagging and Adaptive-Size Hoeffding Tree (ASHT) Bagging [13]. The experimental framework has shown that the new ensemble methods perform very well compared to several known method.

Leverage bagging [14] which improved performance of online bagging algorithm proposed by A. Bifet, etc., it applied randomization to the input and the prediction of outcome in online bagging. Randomization is done in two parts: resampling of data stream increased and error codes are used to detect output. In case of classifiers, Poisson distribution is used to measure occurrences of event in particular block. Leverage bagging method computes the value of Poisson distribution by increasing λ value of resampling. Similarly the output of ensemble classifier is improved by using error-correcting output code method where a multiclass problem is handled by using a binary classifier.

Recently , Carvalho Santos et al. has proposed the Adaptable Diversity-based Online Boosting (ADOB)[15] algorithm, improved version of the online boosting, as concept drift encounters it employees recovery mechanism of expert. In this method, distribution of instances is done more efficiently among experts, by which it adapts to concept drifts more quickly. In case of abrupt concept drift it works well. This is done by reducing and increasing value of λ based on the classification accuracy. ADOB technique improves the accuracy of classifiers, reduced execution time and memory usage.

ADOB is improved by Barros, et. al named as Boosting-like Online Learning Ensemble (BOLE) [16]. All previously designed online boosting methods worked on a basic requirement that each classifier error must be lower than half of prediction output but it seems very hard when the problem is not binary. Similarly this algorithm removes the classes related to the failed classifier, resulting in poor performance. In case of BOLE, there are two different parameters “breakVotes” and “errorBound” which put restriction on classifiers for voting as well as it avoids negative voting by using “weightShift” parameter whose range is 0.0 to 5.0. It also used different concept drift detection method. Thus this method has made improvements by providing voting requirement and adaptation in concept drift detection method.

Frias-Blanco *et al.* [17] developed the Fast Adaptive Stacking of Ensembles (FASE), which works on adaptive learning method. It`s learners adapt after detection of concept change and then update the learning model by estimating that concept drift. It uses Na¨ıve Bayes (NB) or Perceptron as base classifier and meta-learner. It applied drift detection method called as Hoeffding-based Drift Detection Method (HDDM*A*-test) [18] to detect changes means ups and downs in error rates. In this technique each input data stream instance get processed in fixed time and space complexity. FASE uses a test-then-train approach [19] to generate meta-instances. It uses n base adaptive classifiers which provide meta- instances to the Meta adaptive classifier by which it performs accurate prediction. FASE uses one change detector which can remove any adaptive base learner whose performance falls below the error rate. Thus any change in input data stream can be detected precisely. But due to use of Perceptron learning is slow compare to NB so some times its performance is reduced.

Recently Jicheng Shan et al. has proposed a new online active learning ensemble framework (OALEnsemble) to identify the concept drift in the data stream [20]. To capture the gradual drift as well as sudden drift OALEnsemble, applied mixed labeling strategy. Two categories of ensembles are used: a stable classifier which learns from all n blocks and dynamic classifiers which learn from most recently arrived data blocks. Finally ensemble classifier combines the weights from these stable classifiers and dynamic classifiers for prediction of class. Here Base classifiers are created and updated by using multilevel windows. Combination of blocked based and online ensemble classifier is effectively implemented in this method.

IV. DATA STREAM CLASSIFICATION USING BLOCK BASED ENSEMBLE METHODS

The second approach to process data stream is to divide data into blocks, which include instances sequential in time. Sometimes these blocks are non-overlapping, sometimes overlapping. These techniques are suitable for sudden and to some extent to incremental drifts.

Instead of working on entire data, W. N. Street et al. has developed very first blocked based ensemble algorithm popularly known as Streaming Ensemble Algorithm (SEA) [21]. In the SEA method all base ensembles are designed using decision trees constructed by applying Quinlan`s C4.5 algorithm [22]. This method employees 20 to 25 classifiers and experimented with “gated” voting procedure, in which a special classifier C_i was trained to decide whether its corresponding classifier member M_i would correctly classify given data point. It provided distinguished blocks of data for each classifiers, one pass processing of data and use of decision trees algorithm made this approach novel initially but it has limitations in terms of prediction accuracy and memory size that has to be overcome.

Meanwhile H. Wang, et al. [7] has proposed advancement in the chunked based data stream processing known as Accuracy Weighted Ensemble (AWE) algorithm that trains an ensemble of classification models, such as C4.5, RIPPER, naive Bayesian, etc., from block of the data stream. This approach divides the incoming data streams in equal size. The classification error on current training set is used to predict weight of that classifier. In the data stream ensemble accuracy of prediction depends upon selection of efficient classifiers, to identify inaccurate base classifier, instead of discarding classifiers randomly the classifier with higher MSE (Mean Square Error) get discarded. Similarly instance based pruning technique is applied to select top K ensembles which keep the top P classifiers with the highest prediction accuracy on the current training data. There are major drawbacks of system like accuracy of prediction depends upon chunk size, higher the chunk size more the error rate. AWE pruning strategies may remove more base classifiers in case of sudden drifts.

Previously mention Online learning process works on selection of classifier based on its performance, in case of Additive Expert (AddExp) [23] system performance of classifier is measure over currently changing concepts and not on previous. Here the AddExp.D algorithm is used for prediction; it maintains weighted ensembles of predictive models, the ensemble works on weighted vote of all base classifiers. The weights of all base classifiers which misclassify the data block get reduced and if overall prediction was incorrect then new classifier is added. Here the classifiers are termed as experts. To restrict the growth of classifiers/ expert, two pruning methods are used to disqualify non performing expert. In first method there is an upper bound used on number of classifiers/ experts. Addition of new expert crosses that bound then oldest expert has to disqualify. But this pruning technique is impractical as recurring data stream prediction may get hamper due to deletion of old record. In second pruning method if newly added expert exceeds than upper limit then the expert with lowest weight get removed. AddExp introduced new pruning and ensemble algorithm but it has limited by various bounds like number of expert, weighted error etc. AddExp used static weight for ensemble this drawback was overcome by J.Z. Kolter, et al. in [24] by introducing dynamic weighted majority (DWM) algorithm, which activate or deactivate base learners or experts in response to expert performance. In case of misclassification, weight of expert get decreased by fixed value □□□□similarly to remove nonperforming classifier it uses threshold value □ and a controlling parameter P is used to limit the adding and removal of expert. Two types of algorithms are used to implement base classifiers: Incremental Tree Inducer, or ITI [25] and incremental version of naive Bayes. Performance of classification is little improved in case of Naïve Bayes algorithm.

The ensemble based learning on chunks of data or in blocked based method is carried out by various researchers [26],[27],[28],[29],[30]. The Learn++.UDNC proposed by Gregory Ditzler et al [26] that allows the classifier to learn step by step the new concept/classes from odd-even datasets. Adaptation of classifier is weighted then it normalized previous confidence measures and a new transfer function that is used to reduce the confidence bias of a sub-ensemble trained on a majority class. Whereas Magdalena Deckert has developed a framework called Batch Weighted Ensemble BWE [27], which works on sudden and gradual drift by using a drift detector into the evolving environment. Advanced version of Learn++.UDNC is developed by Ryan Elwell as Learn++.NSE [28] algorithm, every time it develops a new model for each batch afterwards it combines these models using a dynamically weighted majority voting. Recently Leandro L. Minku et al. introduced a drift detection method called Diversity for Dealing with Drifts (DDD) [29] which maintains ensembles to deal drift with many diversity levels. It used term diversity for the accuracy of two ensembles on same data set on different time intervals. This algorithm shows better performance compared to previously designed algorithms EDDM [30] and DWM [24].

In spite of chunked based and online data stream processing methods, research is also carried on mixed approach in [20, 31-38]. Combination of blocked based and online processing has given different aspect to handle concept drift problem of data stream classification. Adaptive Classifiers-Ensemble (ACE) employed one online classifier, many batch classifiers and a drift detection mechanism [31]. It used sliding window and long time buffers to retain classifier results for longer time. Pruning method of weakest first is used to remove classifier. Along with this weighting majority is used for ensemble output. All these makes ACE more powerful system. More advances like use of different algorithms, pruning techniques and accuracy improvement were added in subsequent systems to mention few MCIK [32], CCP[33], DXMiner [34], AE [35], DACC [36] and ADACC[37], OAUE[38]. Recently Beygelzimer has developed two boosting algorithms Online BBM.W and AdaBoost.OL.W [39], which takes benefits of online working and sampling concept from AdaBoost [40] technique. Data stream clustering is done in [41] which has used Nearest neighbour approach. Similarly live data streaming is also done in [42] which has applied group learning concept and improved performance of stream clustering. Thus tremendous research is done on data stream classification technique.

V. CONCLUSION

Thus this study focuses on two main categories of data stream processing. Data stream classification has several difficulties, as it involves maintaining an accurate ensemble model which should cope with the speed of data arriving using limited resources. As described in this study, Ensemble based methods are more suitable for classification of data streams. Each advanced method and framework often achieved high accuracy and prediction efficiency. The major issue of different types of concept drifts can be handled by application of various base algorithms. This paper presents the main characteristics of data streams, emphasis on milestone methods for online ensemble and blocked based ensemble is also carried out.

Data stream classification using adaptive ensemble based classifiers will be thoroughly explored for big data stream learning, instead of single ensemble method. More focus will be put on parallel data processing and use of AI may add some interesting features in it. Also, there is a scope for combining traditional classification algorithm to deal with data which is diverse in nature. Similarly anomaly detection and abrupt drift detection is also one area of study. Finally, we can conclude that thought data stream ensemble has strong background of scientific study still there is vast area to explore.

Table 1 describes design and working of some of the remarkable ensemble based online data stream classifiers.

Table 1: Detail description of online ensemble based classifiers

Algorithm	**Design/ Base method**	**Working**
Online Bagging and Boosting N.C. Oza and S. Russell [11,12]	Flat architecture with majority voting scheme for ensemble. Bagging and boosting methods.	First online version of data stream processing. It Bootstrapped data set by sampling of data with replacement.
ADWIN Bagging Bifet, A., Holmes et al. [13]	Incremental and adaptive learning, used window. Hoeffding Tree classifier.	It used Hoeffding Tree of different sizes to detect gradual and abrupt change.
Leverage Bagging Bifet, A., Holmes et al. [14]	Flat architecture with majority voting scheme incremental adaptive learning. Bagging method.	Input randomization is done by increasing resampling and output detection codes are used.
ADOB S. G. T. de Carvalho Santos et al. [15]	Flat architecture with Hoeffding Tree architecture, majority voting scheme.	Ranking of expert is done based on their prediction accuracy. It is efficient in often and abrupt changes.
BOLE R. S. M. de Barros, et al. [16]	Flat architecture. Improved Oza Boost algorithm with drift detecting mechanism.	Test and train method uses 50% continue voting strategy.
FASE I. Frías-Blanco, et al. [17]	Adaptive learners, it changes set of base classifiers. Naïve Bayes and Hoeffding tree.	Uses Naïve Bayes classifier and Hoeffding based Drift Detection method. It uses Perceptron to evaluate performance.
HDMM I. Fr´ıas, J. del Campo, et al. [18]	Adaptive classifier based on Hoeffding inequality.	It works on moving averages and weighted moving averages, based on three levels stable, warning and drift.
OALE Jicheng Shan , Hang Zhang et al.[20]	Online adaptive classifiers, it`s base classifier is Hoeffding Tree.	It uses hybrid labeling strategy and multilevel sliding window to handle gradual and sudden drift.On demand labeling scheme.
DWM Kolter, M.A. Maloof [24]	Works on Naïve Bayes classifier and Incremental Tree Inducer.	It uses weighted sum policy for prediction. Adds and removes expert learners based on their global performance.

REFERENCES

[1] https://searchdatamanagement.techtarget.com/definition/big-data.

[2] Read, J., Bifet, A., Holmes, G., & Pfahringer, B. "Efficient multi-label classification for evolving data streams." Technical Report University of Waikato, Department of Computer Science, 2010.

[3] I. Khan, J.Z. Huang, K. Ivanov. "Incremental density-based ensemble clustering over evolving data streams." Neurocomputing, 191, 34-43, 2016.

[4] I. Zliobaite, A. Bifet, B. Pfahringer, G. Holmes, "Active learning with drifting streaming data", IEEE Transactions on Neural Networks & Learning Systems, 25, 27-39,2014.

[5] https://en.wikipedia.org/wiki/Concept_drift 09-22-2018.

[6] Zliobaite I."Learning under concept drift: an overview." Technical report faculty of mathematics and informatics, Vilnius University, Vilnius, Lithuania, 2009.

[7] Wang H., FanW., Yu P. and Han J. "Mining Concept-Drifting Data Streams using Ensemble Classifiers." in the 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Washington DC, USA, 2003.

[8] D. Brzezinski, J. Stefanowski. "Reacting to different types of concept drift: the accuracy updated ensemble algorithm." IEEE Trans. Neural Netw. Learn. Syst. 25 (1) 81–94, 2014.

[9] D. Brzezinski , J. Stefanowski. "Prequential AUC: properties of the area under the ROC curve for data streams with concept drift." *in* Knowledge and Information Systems 52(2):531-562 · January 2017.

[10] Heitor Murilo Gomes, Jean Paul Barddal, And Fabr´Icio Enembreck."A Survey on Ensemble Learning for Data Stream Classification."ACM Computing Surveys, Vol. 50, No. 2, Article 23, March 2017.

[11] N.C. Oza and S. Russell. "Online bagging and boosting.", Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics (AIS- TATS'01), Morgan Kaufmann, Key West, USA, p. 105-112,2001.

[12] Nikunj C. Oza and Stuart Russell. ''Experimental Comparisons of Online and Batch Versions of Bagging and Boosting." The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, pp. 359-364, August 2001.

[13] Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., Gavald`a, R.: New ensemble methods for evolving data streams. In: KDD, pp. 139–148 2009.

[14] A. Bifet, G. Holmes, and B. Pfahringer, "Leveraging bagging for evolving data streams," in *Proc. ECML/PKDD, I*, pp. 135–150,2010.

[15] S. G. T. de Carvalho Santos, P. M. G. Júnior, G. D. dos Santos Silva, and R. S. M. de Barros, "Speeding up recovery from concept drifts," in *Proc. Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases*, pp. 179–194,2014.

[16] R. S. M. de Barros, S. G. T. de Carvalho Santos, P. M. G. Júnior, "A boosting-like online learning ensemble," in *Proc. IEEE IJCNN*, Jul. 2016, pp. 1871–1878.

[17] I. Frías-Blanco, A. Verdecia-Cabrera, A. Ortiz-Díaz, and A. Carvalho, "Fast adaptive stacking of ensembles," in Proc. ACM Symp. Appl. Comput., 2016, pp. 929–934.

[18] I. Fr´ıas, J. del Campo, G. Ramos, R. Morales, A. Ortiz, and Y. Caballero. "Online and non-parametric drift detection methods based on Hoeffding's bounds." IEEE Transactions on Knowledge and Data Engineering, 14(3):810–823, 2015.

[19] J. Gama, P. Rodrigues, and G. Castillo. "Evaluating Algorithms that Learn from Data Streams." In *Proc. 2009 ACM Symposium on Applied Computing*, pages 1496–1500, 2009.

[20] Jicheng Shan , Hang Zhang, Weike Liu, and Qingbao Liu. "Online Active Learning Ensemble Framework for Drifted Data Streams.", in Proc. IEEE Transactions On Neural Networks And Learning Systems, 2018, pp.1 – 13.

[21] W. N. Street and Y. Kim. "A streaming ensemble algorithm (sea) for large-scale classification." ACM KDD Conference, pages 377–382, 2001.

[22] J. R. Quinlan. C4.5: programs for machine learning. Morgan Kaufmann, San Mateo, CA, 1993.

[23] Jeremy Z. Kolter and Marcus A. Maloof. "Using additive expert ensembles to cope with concept drift." In Proceedings of the 22nd International Conference on Machine Learning (ICML'05). ACM, New York, NY, 2005, 449–456.

[24] Kolter, M.A. Maloof. "Dynamic weighted majority: an ensemble method for drifting concepts." Journal of Machine Learning Research 8 (2007) 2755-2790.

[25] P. E. Utgoff, N. C. Berkman, and J. A. Clouse."Decision tree induction based on efficient tree restructuring." *Machine Learning*, 29:5–44, 1997.

[26] Gregory Ditzler, Michael D. Muhlbaier, and Robi Polikar. Incremental learning of new classes in unbalanced datasets: Learn++.UDNC. In *Multiple Classifier Systems*. Lecture Notes in Computer Science, Vol. 5997. Springer, Berlin, 33–42, 2010.

[27] Magdalena Deckert. "Batch weighted ensemble for mining data streams with concept drift." In Foundations of Intelligent Systems. Springer, 290–299, 2011.

[28] Ryan Elwell and Robi Polikar. "Incremental learning of concept drift in nonstationary environments." IEEE Trans. Neur. Netw. 22, 10 (2011), 1517–1531, 2011.

[29] Leandro L. Minku and Xin Yao. " DDD: A new ensemble approach for dealing with concept drift. " IEEE Trans. Knowl. Data Eng. 24, 4, 2012.

[30] M. Baena-Garcia, J. Del Campo-_ Avila, R. Fidalgo, and A. Bifet, "Early Drift Detection Method," Proc. Fourth ECML PKDD Int. Workshop Knowledge Discovery from Data Streams (IWKDDS '06),pp. 77-86, 2006.

[31] Kyosuke Nishida, Koichiro Yamauchi, and Takashi Omori Ace: Adaptive classifiers-ensemble system for concept-drifting environments. In Multiple Classifier Systems. Springer, 176–185, 2005.

[32] Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham. "A practical approach to classify evolving data streams: Training with limited amount of labeled data." In 8th IEEE International Conference on Data Mining, 2008 (ICDM'08). IEEE, 929–934, 2008.

[33] Ioannis Katakis, Grigorios Tsoumakas, and Ioannis Vlahavas. "Tracking recurring contexts using ensemble classifiers: An application to email filtering." Knowl. Inf. Syst. 22, 3,371–391, 2010.

[34] MohammadM.Masud, Qing Chen, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham. "Classification and novel class detection of data streams in a dynamic feature space." In Machine Learningand Knowledge Discovery in Databases. Springer, 337–352. 2010.

[35] Peng Zhang, Xingquan Zhu, Yong Shi, Li Guo, and XindongWu." Robust ensemble learning for mining noisy data streams." *Dec. Supp. Syst.* 50, 2 469–479, 2011.

[36] Ghazal Jaber, Antoine Cornu´ejols, and Philippe Tarroux. " Online learning: Searching for the best forgetting strategy under concept drift." In Neural Information Processing. Springer, 400–408, 2013.

[37] Ghazal Jaber, Antoine Cornu´ejols, and Philippe Tarroux. A new on-line learning method for coping with recurring concepts: The ADACC system. In *Neural Information Processing*. Springer, 595–604. 2013.

[38] Dariusz Brzezinski and Jerzy Stefanowski. 2014. Combining block-based and online methods in learning ensembles from concept drifting data streams. *Inf. Sci.* 265,50–67, 2014.

[39] Alina Beygelzimer, Satyen Kale, and Haipeng Luo. "Optimal and adaptive algorithms for online boosting." 2323–2331, 2015.

[40] Yoav Freund, Robert E. Schapire," Experiments with a new boosting algorithm", In *ICML*,Vol. 96. 148–156. 1996.

[41] Pratap Shinde; M. D. Ingle, "Streaming Data Clustering using Incremental Affine Propagation Clustering Approach", International Journal of Science and Research (IJSR), **https://www.ijsr.net/search_index_results_paperid.php?id=SUB156911**, Volume 4 Issue 7, July 2015, 2209 – 2212.

[42] Spraha Kamriya; Vandana Kate, "Live Data Stream Classification for Reducing Query Processing Time: Design and Analysis", International Journal of Science and Research (IJSR), **https://www.ijsr.net/search_index_results_paperid.php?id=ART20174017**, Volume 6 Issue 6, June 2017, 1711 – 1716.

AUTHORS PROFILE

Mrs. R C Samant received her Bachelor of Engineering (Computer) from Mumbai University , India in 2000 and Master of Engineering (Computer) from Thadomal Sahani Engineering College , Mumbai University in year 2010. She is currently pursuing Ph.D. from Computer Engineering Department ,Bharati Vidyapeeth Deemed to be University, Pune. She currently working as Assistant Professor in Gokhale Education Society`s R. H. Sapat College of Engineering Management Studies & Research , Nashik since 2013.She has published more than 15 research papers in reputed international journals including IJCA, IJSET, IRJET. Her main research work focuses on Data Mining, Information Retrival, Big Data Analytics, and Plagiarism checking.

Mr D M Thakore received B. E. (Computer) from Shivaji University in 1990, M. E. (computer) from Bharati Vidhyapeeth Deemed University in 2004. He has completed his PhD from JJTU Rajasthan. HE is currently working as Professor and Head, Department of Computer Engineering, Bharati Vidyapeeth Deemed to be University, College of Engineering, Pune. He has published more than 60 research papers in reputed international journals including IEEE, IJSCE, IJCSE, IJSCNS. His main research work focuses on Database Systems, Software Engineering, Big Data Analytics , Data Mining and Networking. He has 26 years of teaching 15 years of research and 3 years of Industrial experience.