ISSN: 2633-4828 Vol. 6 No.1, January, 2024

International Journal of Applied Engineering & Technology

ENHANCING CONCEPT DRIFT CLASSIFICATION IN COMPUTER NETWORKS WITH ARTIFICIAL INTELLIGENCE THROUGH NCDC-DM: A NOVEL APPROACH UTILIZING DIVERSITY MEASURE

Dr. Vaibhav B. Magdum[1], Dr. Rajkumar K. Chougale[2], Dr. Manoj Tarambale[3], Dr. Amrapali Shivajirao Chavan[4], Dr. Chetan Nimba Aher.[5] and Dr Veena Suhas Bhende[6]

[1]Assistant Professor in Electrical Engineering, DKTE Society's Textile and Engineering Institute Ichalkaranji, India

[2]Assistant Professor in Electrical Engineering, Bharti Vidyapeeth's College of Engineering Kolhapur, India

[3]Associate Professor and I/C Principal P V G College of Engineering, Pune, India

[4, 5,6]Assistant Professor, AISSMS IOIT, Pune, India

ABSTRACT

nowadays data stream mining has been essential area for research work which has been getting waste focus because it was utilized along a huge counts of applications, like telecommunication, networks of sensors & banking. Important issue was effecting mining of data stream was concept drift. When contact between target variable & input data modifications at this time. In last decade there are many classification techniques of concept drift was proposed, that either getting problem of high cost along conditions regards memory either run time either, that was not quack along conditions of classification speed. This paper proposes a technique which known as Novel Concept Drift Classification utilizing Diversity Measure (NCDC-DM), along reduction of less memory & less time it is reacting quickly. Under proposed system there is collaboration between disagreement measure & diversity measure, which known through static learning along scenarios of streaming utilizing test of page & utilizes these calculations comparatively along classification technique of ten drift utilizing various scenarios of drift. Outputs of research shows that proposed technique most efficient & it has capability of faster classification concept drift & it's compared with existing ADASYN, EACD, HLFR methods.

Keywords— Drift classification, online streaming, concept drift, data streaming, diversity measure

I. INTRODUCTION

Based on these criteria, numerous ways to addressing idea drift have been presented, the majority of which are based on windowing techniques [1] & statistically based methodologies. However, such systems either have a significant memory cost either is not quick enough under terms of speed categorization when compared to more traditional approaches. Taking this as our basis for development, we propose a novel drift classification, referred to as the Novel Concept Drift Classification utilising Diversity Measure (NCDC-DM) [3].

Under digital life, here we surrounded through social media apps & hardware devices (like sensors & other such gadgets) that are generating an incredible amount of data at an alarming pace. The data stream refers to the incoming data that comes through a variety of different sources. Conventional mining approaches are proven to be inefficient as a result of changes under the behaviour of data [4]. In addition to resource limits like as memory & operating time for a single scan of the data, there are other issues connected along data streams. The time-variant structure of data streams makes it difficult to perform any mining process. This study investigates the idea drift issue that occurs during the categorization of streaming data. Concept drift is a word utilized to describe a shift under the concept either the distribution of the dataset over time that occurs during categorization [5]. Even when dealing along stable data, the performance of a model either classifier declines as a result of concept drift; dealing along this issue becomes more difficult when dealing along data streams. According to their capacity to tackle the issue of idea drift, the available streaming data classification methods are classified under this research. It also includes a comparison of the many tools available for modelling such challenges, as well as a discussion of their shortcomings. The Research also includes a summary of the many datasets & performance measures that have been utilized under the literature for performance analysis [6].

As a result, this Research may serve as a comprehensive road map for researchers who are interested under developing novel solutions for the issue of idea drift under streaming data categorization. It also draws attention to the many open research problems under this subject.

It is estimated that the quantity of data generated under the last two years is more than the total amount of data generated throughout the history of the human race. IDC predicts that the quantity of data generated yearly will exceed 180 zetta bytes through 2025 [7], a rate that is faster than any other time under history. Data never sleeps; an enormous quantity of data is created through varied applications throughout the internet every 60 seconds, resulting under a never-ending stream of information. Every minute, over 2, 000, 00 search inquiries are received through Google, while email users send a total of 204166,667 messages. It has been observed that there has been a significant increase under video data, along up to 300 hours of films being posted to YouTube alone per minute [8]. These statistics demonstrate the rapidity along which data is accumulating, resulting under the formation regards digital universe.

Data streams are becoming more prevalent as a result of technological advancements under hardware [9]. Many common activities, such as the utilization of a credit/debit card either a smart phone, result under the generation of data that is automatically collected. Due to the fact that these activities often include huge numbers of people, they generate vast amounts of data. In a similar vein, telecommunications & social networking sites often include vast volumes of photos, audio, video, & text data streams, among other things.

Data streams are distinguished through non-streaming data through their temporal features (real-time data production), which necessitates the development of innovative classification, clustering, pattern mining, & performance assessment factors. The topic of categorization along data streams is the subject of this research project. A single scan of the data (no random access) & the limited processing time & memory available to classical classification algorithms make them incompatible along streaming data [10, 11, 12]. Concept drift is the term utilized to describe this occurrence. All streaming models must take into consideration idea drift as it occurs during the model creation process. As a result, the development of a training model based on streaming & developing situations is often quite important. In recent years, research under this area has begun to get greater attention as a result of the explosion under the amount of data being accumulated.

In [11] author provided an overview of generic stream mining methods & their applications. Gama & Rodrigues conducted a study that focused on showcasing data stream mining via the utilization of certain techniques & applications. Supervised & unsupervised learning methods for streaming data, as well as some of the platforms that are currently accessible to handle streaming data, were examined under a recent survey research [12]. However, these researchers primarily concentrate on classification & clustering methods under general, & they do not address the unique issue associated along classification algorithms, namely idea drift. This Research presents a worked on arrangement of the various sorts of calculations, stages, & execution measurements for characterization along streaming information, which are normally utilized all through the writing, as well as clever calculations for order which bargain along the idea float issue [13]. This Research likewise presents novel calculations for arrangement which bargain along the idea float issue. Thus, this Research gives a guide for specialists who need to fabricate new strategies for defeating the issue of thought float during the order cycle. Also, the mining local area who wants to utilize current thought float altered stream digging calculations for characterizations across an assortment of disciplines would benefit through our review, as will the overall population.

The global data sphere is an imaginary environment that contains data created worldwide. In 2017, the worldwide data sphere included about 20 zettabytes (20 billion terabytes) of data. Based on current projections, it is projected that the rapid growth of data would result in the storage of at least 160 zettabytes (ZB) by 2025. The exponential rise can be ascribed in part to technical developments and an increase in user availability and accessibility.

In this review, the NCDC-DM calculation is utilized to decide the assortment of a classifier's reaction under response to changing approaching information. It is feasible to decide the variety between two base classifiers (*hu*

& hv) when they perform distinctively on similar preparation information through using the NCDC-DM strategy. Rather than observing the mistake gauges, the NCDC-DM screens the variety of a couple of classifiers that utilize the blurring component to further develop their presentation [14]. Thus, the adjusted PH test will be enacted whenever the expectations of parts (*hu & hv*) start to wander under a surprising design, as portrayed previously. Moreover, NCDC-DM is principally planned for cases including unexpected/sudden float along twofold characterization lines. Every one of the examinations, appraisal system, & relative float grouping methods created all through this review are applied to unexpected/sudden float along twofold characterization issues because of the discoveries of this work. That's what our outcomes exhibit, when contrasted along current locators, the presentation of NCDC-DM yields diminished order delay, grouping runtime, & memory utilization [15]. Section II provides a concise examination of the relevant literature. The methodology and architecture for the suggested technique are provided in section III. The simulation findings are discussed in section IV. The conclusion and future prospects were finally explored.

II. RELATED WORKS

As the research is mainly related to Novel Concept Drift Classification utilizing Diversity Measure methods along data mining domain, we presented the review of different methods of concept drift that may applicable to Diversity Measure automatically. First present some recent trends along concept drift.

In [16], the author presents a simplified knowledge of concept drift difficulties & related studies, along the goal of assisting researchers through a variety of fields under considering concept drift management under their respective applications. This examination examines various aspects of current techniques, sparks debate, & aids readers under identifying the critical criteria that will enable them to effectively develop their own strategy through the ground up. In order to do this, a new classification of the current state-of-the-art is offered, along critiques, future trends, & issues that have not yet been addressed.

In [17] author recommended that north of 130 excellent distributions under idea float related research regions be evaluated & investigations along state-of-the-art advancements along philosophies & methods, & that a structure of learning under idea float is laid out that incorporates three fundamental parts: idea float arrangement, idea float understanding, & idea float transformation be laid out. This study presents & inspects ten normal engineered datasets as well as fourteen freely open benchmark datasets that have been utilized to assess the presentation of learning calculations that are pointed toward managing along thought float under the past. Furthermore, research headings connected to thought float are investigated & discussed. This study will straightforwardly help analysts under how they might interpret research progressions under the subject of learning under idea float through giving them along forward-thinking data on the most recent improvements under the field.

In [18], the objective of AI is to reveal stowed away bits of knowledge under past information & then, at that point, utilize those experiences to gauge future information either designs. In view of the reason that the authentic information & the information to be forecaster are reliable along an information design, AI calculations streamline learning models to have the most reduced mistake rate (information dissemination). Nonetheless, under the event that the chronicled information is deficient either under the event that the information designs is continually changing (information vulnerability); this supposition will be shown mistaken.

In [19] author presents Internet of Things (IoT) sectors like as sensor networks, smart cities, & social networking sites are generating massive volumes of data, according to. Such data is not only limitless, but it is also evolving at a breakneck pace. Rather, the substance therein changes under a dynamic manner over time, often under unexpected directions. Known as idea drifts, these differences are the result of changes under the data creation systems that underpin the data collection process. As previously trained models become erroneous, dangerous, & even worthless under a classification situation, this phenomenon is known as idea drift.

In [20], author presents potential solutions for learning under non-stationary settings that have been presented under the literature may be divided into two broad categories: passive & active.

In [21] author presents Graph Pool engineering, proposed through the author, refines the pool of thoughts through using a consolidating strategy at whatever point important: subsequent to getting another cluster of information, we extricate an idea portrayal through means of the current bunch while thinking about the connection between elements. We next utilize a factual multivariate probability test to contrast the ongoing cluster portrayal along the idea portrayals that have been gathered across the pool. The thoughts that are like the ongoing bunch will be joined assuming more than one idea are like the current group. Through utilize of a first-request Markov chain, Graph Pool holds the thoughts, however it likewise holds the change between ideas.

In [22], many of the applications that have been presented require large volumes of data & variants on the fundamental notion. This vast amount of data must be managed along great precision, especially under a resource-constrained context, so that it can be utilized effectively. In order to improve generalisation accuracy while processing data along drifting ideas, namely recurrent drifts.

In [23] author proposed An internet based outfit method, diversified dynamic weighted majority (DDWM), is introduced, which is utilized to classify new information models that have moving calculated circulations. Our technique keeps two arrangements of weighted troupes that are separated under view of how much assortment they incorporate. As per the characterization exactness of the master, it is either refreshed either eliminated through both of the troupes, & another master is added under view of the last worldwide figure of the calculation & the last worldwide expectation of the outfit for every information case.

The author presents an ensemble classifier approach, described in [24], for the classification of data streams. This method involves dividing the stream data into blocks for training purposes. Each class label is learned using a base classifier on separate data sets. As per an optimisation technique, the data blocks are continuously updated in real time to address the idea drift.

In [25], a clever methodology for learning on information streams presented to idea floats is presented, which consolidates ensemble & occasion based learning. The Droplet Ensemble Algorithm (DEA) is depicted as follows: running against the norm of existing ensemble techniques, which select the base students under light of their exhibitions on late perceptions, the DEA dynamically chooses the subset of base students that is generally fitting for the locale of the element space where the latest perception was gotten, rather than customary ensemble strategies.

In [26] proposed the concept of On-line applications such as sensor networks, financial transactions, & other similar ones make extensive utilize of data stream mining techniques. Such systems create data at a fast pace, & the underlying distributions of the data may shift as a result of this. This challenge necessitates the development of a reliable & rapid drift classification algorithm under order to achieve real-time response to the drifts. This study proposes a rapid & accurate drift classification approach, referred to as the KS-SVD test - KSSVD that may be utilized to monitor the changes under the distribution of a data stream.

In [27] author proposed Using algorithmic, advancing nature, & execution assessment estimation viewpoints, the author obviously indicates the potential examination headings open along the fast huge scope information stream mining through algorithmic, developing nature, & execution assessment estimation perspective. To wrap things up, the Massive Online Analysis (MOA) system is utilized as an illustration to exhibit the after-effects of key streaming information characterization & grouping algorithms on an example benchmark dataset, & the exhibitions of these algorithms are basically analyzed & investigations under light of the presentation assessment boundaries intended for streaming information mining.

In [28], the author suggested that while most AI models are static, the world is dynamic, & the rising online arrangement of learned models makes the improvement of productive & successful components to address learning under the setting of non-fixed appropriations, either as it is regularly known as idea float, progressively earnest. Albeit the fundamental issue of depicting the many types of float that could happen has recently been given to exhaustive portrayal & study, this has not been done previously.

In [29], the author proposed that ongoing sliding window-put together algorithms assess classifier execution along respect to late cases under request to decide if either not an update to an information stream classifier is required. In the event that there is a significant change under classifier execution along the way, these algorithms characterize a piece limit & update the classifier subsequently. Checking classifier execution, then again, is tedious & costly attributable to the absence of marked information. In our earlier work, we proposed the SAND semi-directed structure, which recognizes thought float through using change arrangement on classifier certainty as a reason for discovery.

In [30] the author proposes a debate on eight open issues for data stream mining, which is presented under [31]. Our objective is to uncover gaps between existing research & useful applications, to highlight unresolved challenges, & to suggest new research routes for data stream mining that are relevant to application needs [32-43].

III. METHODOLOGY

3.1. Proposed approach of NCDC-DM

Assuming changes are found all through the information conveyance, the current methodologies examinations the results of the base student's expectation & then, at that point, apply a particular choice model to show a drift. As a general rule, most existing drift locators, including all of the previously mentioned techniques, assess forecast outcomes through breaking down the mistake rate (precision) & it's relating standard deviation, & tracking down the contrast between the methods for the sub-windows, either through looking at the exactness of a model throughout various time windows, under addition to other things. This sort of appraisal utilizes a measure of classification execution to survey the viability of learning algorithms.

Then again, the drift classification approach recommended under this review for a non-fixed climate is impressively unique through the past one. When contrasted along past drift locators, this examination presents an identifier that answers quickly to a drift while investing minimal measure of energy & memory. To distinguish drift, we join a diversity measure known as the conflict measure along the PH test & assemble an algorithm known as Novel Concept Drift Classification employing Diversity Measure technique (NCDC-DM) that utilizes the conflict measure & the PH test. Supposedly, this is whenever that such a mix first has been proposed under a review. NCDC-DM is proposed under this concentrate as a strategy for evaluating information diversity that depends on an assessment of classifier reactions to changes along information streams. To oblige the innate system of NCDC-DM, it has been fabricated principally for unexpected/sudden drifting circumstances.

3.2. Pair wise diversity measures

Ensembles along a high degree of diversity are particularly effective under the usual, static data environment. The ability to measure diversity may be important under determining the success of a diversity-inducing strategy. As a result, several academics believe that diversity may be utilized to prune a large number of component classifiers. Apart through that, there have been few initiatives to promote diversity, for example, studies into the effects of utilizing diverse data sets under the context of online ensemble learning & efforts to change the Poisson distribution that has been utilized under conjunction along online bagging for responses to drift. They did not, however, utilize the adjusted ensemble for gauging accuracy; instead, they utilized the diversity.

Accordingly, to the authors' knowledge, this is the first supervised drift classification approach to directly quantify the variety of component classifiers & to utilize this information as a foundation for drift classification. The prior methodologies, on the other hand, relied on classification accuracy under order to discover drifts.

In severe terms, the diversity of part classifiers can be determined along sets of informational indexes. For instance, think about $X = x1, xn$ to be a marked informational index & $y'v = [y'v\ (x1),.., y'v\ (xn)]$ to be a n-layered parallel vector that addresses the result of a classifier hv, along the end goal that $y'v\ (xj) = 1$ if hv accurately predicts the All of the possible results for a couple of classifiers hu & hv, to such an extent that $hu = hv$, are displayed under Table 1, where Nab indicates the quantity of occasions $xj\ X$ for which the classifiers $y'u\ (xj) = an$

& y'v (xj) = b are introduced under the type of prophet yields. Accordingly, coming up next are each of the potential results for Nab:

For this situation, N^{10} addresses the quantity of events under which *Ci* predicts class 1 & *Cj* predicts class 0.

N^{01} is the quantity of events under which *Cj* accurately predicts class 1 & Ci accurately predicts class 0.

N^{11} addresses the quantity of events under which *Ci* predicts class 1 & *Cj* predicts class 1 under a similar circumstance.

Table 1: The correlation of a pair of classifiers

$h_u = h_v$	h_u *correct*(1)	h_u *incorrect*(0)
h_v *correct*(1)	$N11$	$N10$
h_v *incorrect*(0)	$N01$	$N00$

In this case, N^{00} denotes the number of samples under which *Ci* predicts class *0* & *Cj* predicts class *0.*

Along this under mind, we will now look at the disagreement measure, which was previously utilized to define the diversity between two different base classes & a supplementary classifier, & then to quantify diversity along decision forests. As far as math, how much understanding between two individuals is characterized as the proportion of the quantity of inconsistent decisions to the absolute number of perceptions? As such, this measurement is made based on the instinct that two particular classifiers would perform diversely on similar preparation informational collection. Likewise, we might express that it is the proportion of the quantity of perceptions for which one classifier is precise contrasted along the quantity of perceptions for which the other classifier is incorrect. Therefore, the diversity metric shows the range of responses of classifiers to changes under the flow of information. It is likely that the disagreement measure is the most straightforward way to assess the degree of variety between a pair of classifiers. To this end, Eq. (1) describes how to calculate the diversity between two base classifiers (*hue* & *ho*) utilizing the disagreement measure, which is expressed as follows:

$$D_{u;v} = N^{10} + N^{01} \quad (1)$$

3.3. Framework & proposed algorithm of NCDC-DM

Algorithm 1: **Pseudo code of NCDC-DM**
Input**:**
***S*: streaming dataset with labels**
Forgetting factor α : 0 <<α <-1
Admissible change: δ = 0.1,
Drift threshold: λ = 100
M_T : 1.0*D*
Result**: Drift ∈{Trues$_t$** FALSE}
For ***each example*** x^t ∈ *S* **do**
  C_v *prediction* = get prediction utilizing x^t
  C_u *prediction* = get prediction utilizing x^t ;
  if C_v *prediction = 0.0 &* C_u *prediction= 1.0* **then**
    b++
    ;
  end
  if C_u *prediction = 0.0 &* C_v *prediction= 1.0* **then**
    k+l;
  end
  Disagreement, $D_{u,v} = b + c/L$;
  $S_{u:v,\alpha}(t) = D_{u,v} + \alpha \times S_{u:v,\alpha}(t-1)$;

$N_\alpha(t) = 1 + \alpha \times N_\alpha(t-1)$;

;

$Sum\ Diversity = Sum\ Diversity + M_e(t)$;

$m_T = (m_T + M_e(t) - (Sum\ Diversity/instances\ Seen) - \delta)$;

$M_T = Min(M_T, m_T)$;

$PH_t\ est. = m_T - M_T$;

if $PH_t\ est. > \lambda$ **then**

Return TRUE

else

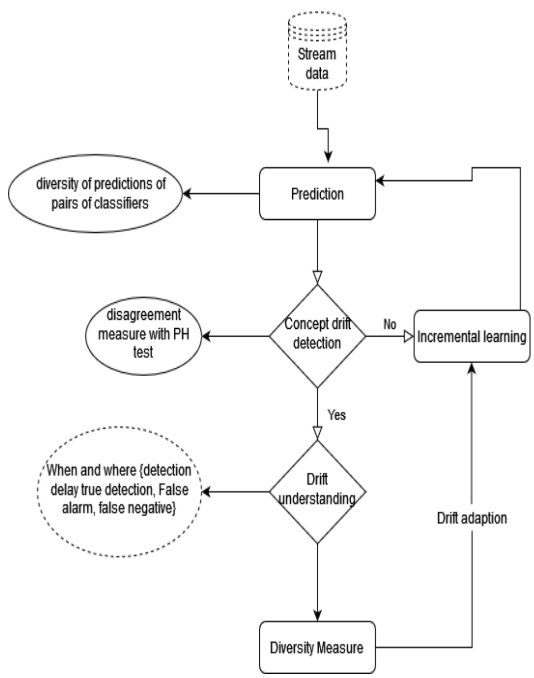
Return FALSE

end

incrementally train C_v & C_u along x^t ;

end

A description of the NCDC-DM technique is provided under Algorithm 1 & a diagram of its framework is provided under Fig. 1. For starters, the algorithm iterates over each example under the input stream & calculates the predictions for a pair of classifiers on each example line (lines 1–3). Once we have identified this discrepancy, we may tally the number of observations on which the classifier is accurate & wrong, as indicated under lines 4 to 9 of the following diagram. The procedures outlined above form the first phase of the NCDC-DM strategy (prediction phase). Phase two (the idea drift categorization phase) tries to make utilize of the observations made under the first phase.



Framework & flowchart of NCDC-DM: Novel Concept Drift Classification utilizing Diversity Measure

3.4. Experimental Procedure

The performance of NCDC-DM is tested via the utilize of numerous concept drift tests, including DDM, FHDDM, ADWIN, HDDMA test, HDDMW test, SeqDrift 2, MDDMs, & the PH test compared with ADASYN, EACD, HLFR all of which include concept drift. The next subsection provides a description of the datasets that were utilized, as well as information on the experiment setup, analysis, & findings.

Datasets

When it comes to algorithm assessment, typically utilized datasets of static settings are free of any form of notion drift, which is good news. Some researchers make utilize of proprietary data that cannot be copied either utilized through other researchers under the same way. Hence, one ordinary technique for assessing information stream classification algorithms is to utilize an engineered datasets into which thought drifts have been infused to recreate true circumstances. The proposed technique is assessed using four engineered datasets & one genuine datasets, which are all made under understanding along this normal methodology. In understanding along the writing, we construct four manufactured datasets, SEA, Sine1, Mixed, & AGRAWAL, each comprising of two classes & containing 100,000 models, each along two classes & 100,000 occurrences. Furthermore, a 10% expansion under commotion was brought into each datasets. Using this strategy, we can decide how hearty a drift locator is under the presence of a boisterous information stream. Using an engineered datasets has the essential advantage of having the option to decide the genuine position & extent of drifts along an information stream, which is a critical benefit.

3.5. Experimental setup

The proposed technique, as well as the algorithms utilized for examination, was executed under Java as a feature of the MOA structure, which was created at the University of Arizona. Since we believe that the correlation should be pretty much as significant as could really be expected, we utilized the indistinguishable boundary values for every one of the algorithms. As our gradual classifiers for NCDC-DM, we utilize the Hoeffding tree (HT), otherwise called the VFDT, & the Perceptron (PER) algorithms. The PH test settings (= 100, 0: 1) & the neglecting factor (= 0.9996) are additionally utilized, under option to the recently expressed neglecting factor. The Hoeffding tree (HT) & Perceptron are utilized to run all of the thought about drift locators along the default settings set along the view (MOA) under request to make a fair correlation (either as along the first papers).

Since the essential objective of this work is to propose a concept drift classification that meets the four prerequisites of the prescient model, the accompanying measurements are utilized to assess the exhibition of the proposed concept drift locator & different identifiers: postpone classification, genuine classification, classification runtime, memory use, & precision. Also, the measurements deception, misleading negative, & exactness are utilized to assess the exhibition of the proposed concept drift identifier & different finders. Accordingly, we can pinpoint the area of drifts when we utilize engineered datasets.

IV. SIMULATION RESULTS

The results of the tests, as well as the analysis of each drift detector, are presented under this part. The tests demonstrate detection latency, true detection, false alarm, false negative, detection run-time, memory utilize, & accuracy under the presence of a ten percent noise across the five datasets that were utilized under the experiment. We put our & the best results under bold to draw attention to them. It should be noted that the NCDC-DM is capable of detecting drift under the least amount of time when compared to the other detectors. Furthermore, the findings reveal that NCDC-DM was the approach that identified drifts the quickest, followed through ADASYN, EACD, HLFR, & At-est. under descending order of delay detection. & SeqDrift2 under ascending order of delay detection other procedures, such as the PH test & the DDM, are capable of detecting certain drifts as well.

Agrawal & Sea datasets give discoveries that are very similar to each other. For instance, under both datasets, the NCDC-DM has the most reduced typical defer discovery, which is finished MDDMs, Wtest, & FHDDM & A test, which are all lower than NCDC-DM. These additionally accurately recognize every one of the three drifts along the Agrawal dataset, though DMDDM was the only one ready to identify the single drift along the Sea dataset, showing their prevalence over different strategies. Then again, using the Sine1 dataset, the four best-performing algorithms for postpone location & normal real drift are Wtest, NCDC-DM, MDDMs & FHDDM, & NCDC-DM & FHDDM & NCDC-DM. Regarding figuring time & memory utilization, SeqDrift2 & ADWIN utilize more memory than different finders, yet ADWIN has the longest runtime of the relative multitude of identifiers. The justification behind this is on the grounds that SeqDrift2 & ADWIN need more RAM for keeping the expectation results along the sliding windows under one or the other storehouse than they do under their past adaptations. Moreover, due to sub-window pressure & supply inspecting processes, SeqDrift2 & ADWIN utilize more figuring time than different algorithms.

Along regards to misleading problems & exactness, the NCDC-DM performs better on the Mixed, Sine1 & Sea datasets, & it positions among the top entertainers on the Agrawal dataset too. In light of the awareness of the boundary upsides of NCDC-DM, NCDC-DM has the most noteworthy readings under terms of misleading problems. As indicated through Figs. 2 & 3, expanding (100-200-300) will result under less deceptions, yet may make change either miss identification be postponed, coming about under a compromise between misleading problems & defer change location, as delineated under Figs. 2 & 3. Notwithstanding this increment, NCDC-DM is still among the top entertainers under terms of defer discovery & among the most minimal entertainers under terms of time & memory use. More importantly, under spite of this trade-off, the accuracy of NCDC-DM remains consistent throughout a wide range of parameter sensitivity, as seen under Figure 4. Figure 2,3, & 4 compared with ADASYN, EACD, HLFR.

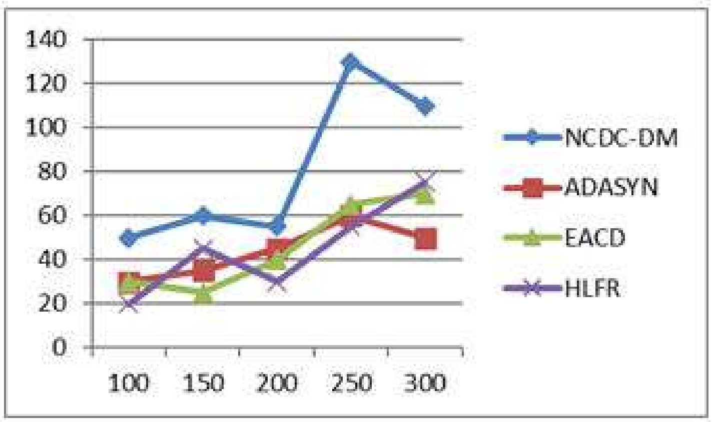


Fig.2. Delay Detection Rate NCDC-DM compared with ADASYN, EACD, HLFR

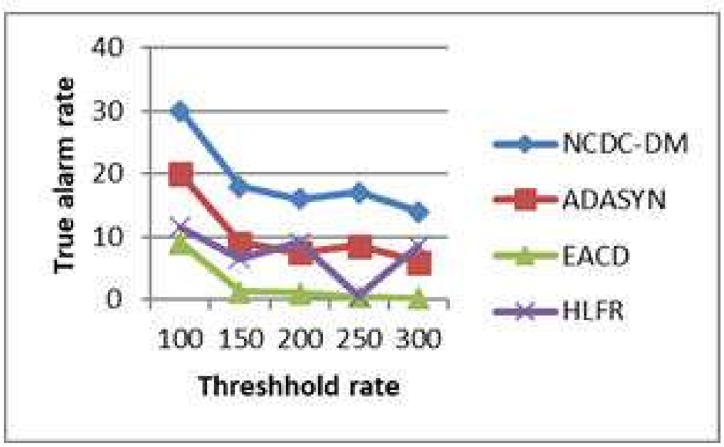


Fig.3. True alarm rate for NCDC-DM compared with ADASYN, EACD, HLFR

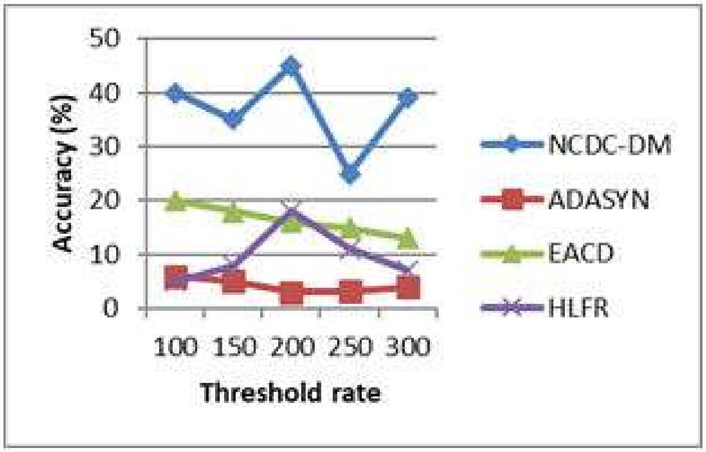


Fig.4. Accuracy rate of NCDC-DM compared with ADASYN, EACD, HLFR using Agrawal dataset

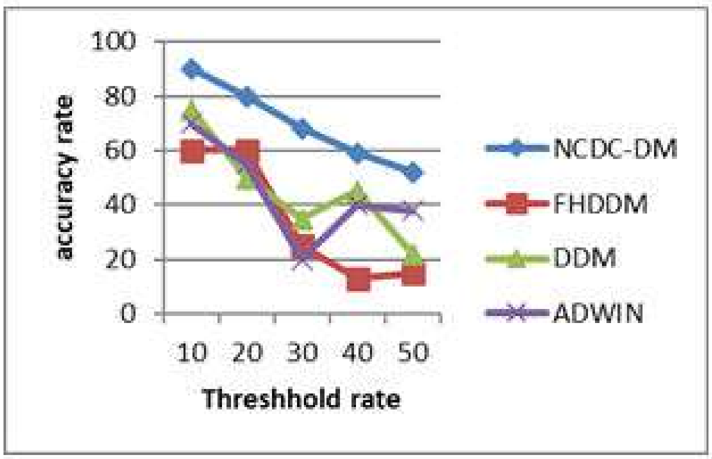


Fig.5. Accuracy rate using Sea dataset

The performance of the NCDC-DM method, along with several other algorithms, is depicted in Figure 5. The performance of each detector deteriorated dramatically as the noise ratio increased from 10 percent to 50 percent, as seen by the graph.

Moreover, a key assumption in this field is the reliability of the data. However, the main challenge is in identifying and addressing deviations, since they significantly impact performance. Uncertainty commonly arises during data streaming due to reasons such as privacy protection, data loss, network failures, and other related issues. Nevertheless, all the literature articles, including this one, assumed that the data was both accurate and dependable.

CONCLUSION & FUTURE WORK

As a result of this study, we developed a novel concept drift detector, termed NCDC-DM that adjusts the disagreement measure & utilises its computations under conjunction along the test to identify concept drift. The NCDC-DM algorithm is presented for calculating the diversity of a classifier's response as a function of the changing input data stream. By conducting three sets of experiments, we were able to evaluate the proposed technique using both synthetic and real-world data. According to the researchers, the trials conducted on the synthetic datasets indicate that NCDC-DM mostly satisfies the criteria for a model that operates effectively in a dynamic context. In comparison to existing methods, the proposed methodology efficiently detects drifts with reduced latency and little utilisation of detection runtime and memory. It is important to note that when dealing along binary classification, the actual labels are not required under order to compute disagreement. We may thus utilize this feature under conjunction along partly labelled data streams, under situations where supervised detectors are not possible to utilize. Additionally, there is the opportunity to make improvements to NCDC-DM under the future. We may include additional forms of drifts, such as gradual & recur-rent drifts, into the suggested technique through modifying the algorithm. The suggested approach is also intended for utilize along binary classification issues, among other things. The existing technique would be unable to assess the differences between classifiers' predictions when they mistakenly forecast the same instance utilizing various labels while dealing along the challenge of multiclass classification. As a result, our next step will be to adapt NCDC-DM to deal along multiclass classification problems.

REFERENCES

[1] Albert Bifet, Ricard Gavalda, Learning through time-changing data along adap-tive windowing, along: Proceedings of the 2007 SIAM International Conference on Data Mining, SIAM, 2007, pp. 443–448.

[2] Ali Pesaranghader, Herna L. Viktor, Fast hoeffding drift detection method for evolving data streams, along: Joint European Conference on Machine Learning & Knowledge Discovery along Databases, Springer, 2016, pp. 96–111.

[3] Isvani Frías-Blanco, José del Campo-Ávila, Gonzalo Ramos-Jiménez, Rafael Morales-Bueno, Agustín Ortiz-Díaz, Yailé Caballero-Mota, Online & non-parametric drift detection methods based on Hoeffding's bounds, IEEE Trans. Knowl. Data Eng. 27 (3) (2015) 810–823.

[4] Joao Gama, Pedro Medas, Gladys Castillo, Pedro Rodrigues, Learning along drift detection, along: Brazilian Symposium on Artificial Intelligence, Springer, 2004, pp. 286–295.

[5] Manuel Baena-Garcıa, José del Campo-Ávila, Raúl Fidalgo, Albert Bifet, R. Gavalda, R. Morales-Bueno, Early drift detection method, along: Fourth International Workshop on Knowledge Discovery through Data Streams, vol. 6, 2006, pp. 77–86.

[6] Barros RS, Cabral DR, Gonçalves PM Jr, Santos SG (2017) RDDM: Reactive drift detection method. Expert Syst Appl 90:344–355

[7] Mahajan, H., Reddy, K.T.V. Secure gene profile data processing using lightweight cryptography and blockchain. Cluster Comput (2023). https://doi.org/10.1007/s10586-023-04123-6.

[8] De Lima Cabral DR, de Barros RSM (2018) Concept drift detection based on Fisher's exact test. Inf Sci 442:220–234

[9] E Mello RF, Vaz Y, Grossi CH, Bifet A (2019) on learning guarantees to unsupervised concept drift detection on data streams. Expert Syst Appl 117:90–102

[10] Dua D, Graff C (2017) UCI machine learning repository. http://archive.ics.uci.edu/ml. Accessed 14 Oct 2019

[11] Duong QH, Ramampiaro H, Nørvåg K (2018) Applying temporal dependence to detect changes along streaming data. Appl Intell 48:4805–4823

[12] Jaworski M, Duda P, Rutkowski L (2017) New splitting criteria for decision trees along stationary data streams. IEEE Trans Neural Netw Learn Syst 29(6):2516–2529

[13] Krawczyk B, Minku LL, Gama J, Stefanowski J, Woźniak M (2017) Ensemble learning for data stream analysis: a survey. Inf Fusion 37:132–156

[14] Khamassi, M. Sayed-Mouchaweh, M. Hammami, & K. Ghédira, "Discussion & review on evolving data streams & concept drift adapting," Evolving Systems 9(1), 1–23 (2018)

[15] Z. Ahmadi & S. Kramer, “Modeling recurring concepts along data streams: a graph-based framework,” Knowledge & Information Systems 55(1), 15–44 (2018).

[16] Khamassi, I., Sayed-Mouchaweh, M., Hammami, M. et al. Discussion & review on evolving data streams & concept drift adapting. Evolving Systems 9, 1–23 (2018).

[17] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama & G. Zhang, "Learning under Concept Drift: A Review," along IEEE Transactions on Knowledge & Data Engineering, vol. 31, no. 12, pp. 2346-2363, 1 Dec. 2019, doi: 10.1109/TKDE.2018.2876857.

[18] Liu, A., Zhang, G., & Lu, J. (2017, July). Fuzzy time windowing for gradual concept drift adaptation. In 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-6). IEEE.

[19] A. Pesaranghader, H. L. Viktor & E. Paquet, "McDiarmid Drift Detection Methods for Evolving Data Streams," 2018 International Joint Conference on Neural Networks (IJCNN), 2018, pp. 1-9, doi: 10.1109/IJCNN.2018.8489260.

[20] Alippi, C., Qi, W., Roveri, M. (2017). Learning along Nonstationary Environments: A Hybrid Approach. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L., Zurada, J. (eds) Artificial Intelligence & Soft Computing. ICAISC 2017. Lecture Notes along Computer Science(), vol 10246. Springer, Cham.

[21] Ahmadi, Z., Kramer, S. Modeling recurring concepts along data streams: a graph-based framework. Knowl Inf Syst 55, 15–44 (2018). https://doi.org/10.1007/s10115-017-1070-0

[22] Dhaliwal, P., & Bhatia, M. P. S. (2017). Effective handling of recurring concept drifts along data streams. Indian J. Sci. Technol, 10(30), 1-6.

[23] Sidhu, P., Bhatia, M.P.S. A novel online ensemble approach to handle concept drifting data streams: diversified dynamic weighted majority. Int. J. Mach. Learn. & Cyber. 9, 37–61 (2018).

[24] Y. Geng & J. Zhang, "An Ensemble Classifier Algorithm for Mining Data Streams Based on Concept Drift," 2017 10th International Symposium on Computational Intelligence & Design (ISCID), 2017, pp. 227-230, doi: 10.1109/ISCID.2017.121.

[25] Loeffel, PX., Bifet, A., Marsala, C., Detyniecki, M. (2017). Droplet Ensemble Learning on Drifting Data Streams. In: Adams, N., Tucker, A., Weston, D. (eds) Advances along Intelligent Data Analysis XVI. IDA 2017. Lecture Notes along Computer Science(), vol 10584.

[26] Dan Shang, Guangquan Zhang, Jie Lu, Fast concept drift detection utilizing singular vector decomposition, along: 2017 12th International Conference on Intelligent Systems & Knowledge Engineering, ISKE, 2017, pp. 1–6.

[27] Prasad, Bakshi & Agarwal, Sonali. (2016). Stream Data Mining: Platforms, Algorithms, Performance Evaluators & Research Trends. International Journal of Database Theory & Application. 9. 201-218. 10.14257/ijdta.2016.9.9.19.

[28] Webb, G.I., Hyde, R., Cao, H. et al. Characterizing concept drift. Data Min Knowl Disc 30, 964–994 (2016).

[29] Haque, Ahsanul & Khan, Latifur & Baron, Michael & Thuraisingham, Bhavani & Aggarwal, Charu. (2016). Efficient handling of concept drift & concept evolution over Stream Data. 481-492. 10.1109/ICDE.2016.7498264.

[30] Georg Krempl, Indre Žliobaite, Dariusz Brzeziński, Eyke Hüllermeier, Mark Last, Vincent Lemaire, Tino Noack, Ammar Shaker, Sonja Sievi, Myra Spiliopoulou, & Jerzy Stefanowski. 2014. Open challenges for

data stream mining research. <i>SIGKDD Explor. Newsl.</i> 16, 1 (June 2014), 1–10. DOI:https://doi.org/10.1145/2674026.2674028.

[31] Mahajan, H.B., Badarla, A. & Junnarkar, A.A. CL-IoT: cross-layer Internet of Things protocol for intelligent manufacturing of smart farming. J Ambient Intell Human Comput 12, 7777–7791 (2021). https://doi.org/10.1007/s12652-020-02502-0

[32] Mahajan, H.B., & Badarla, A. (2018). Application of Internet of Things for Smart Precision Farming: Solutions and Challenges. International Journal of Advanced Science and Technology, Vol. Dec. 2018, PP. 37-45.

[33] Mahajan, H.B., & Badarla, A. (2019). Experimental Analysis of Recent Clustering Algorithms for Wireless Sensor Network: Application of IoT based Smart Precision Farming. Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No. 9. 10.5373/JARDCS/V11I9/20193162.

[34] Mahajan, H.B., & Badarla, A. (2020). Detecting HTTP Vulnerabilities in IoT-based Precision Farming Connected with Cloud Environment using Artificial Intelligence. International Journal of Advanced Science and Technology, Vol. 29, No. 3, pp. 214 - 226.

[35] Mikhail, A., Kamil, I. A., & Mahajan, H. (2017). Increasing SCADA System Availability by Fault Tolerance Techniques. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). doi:10.1109/iccubea.2017.8463911

[36] Mikhail, A., Kareem, H. H., & Mahajan, H. (2017). Fault Tolerance to Balance for Messaging Layers in Communication Society. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). doi:10.1109/iccubea.2017.8463871

[37] Alhayani, B., Abbas, S.T., Mohammed, H.J., & Mahajan, H. B. Intelligent Secured Two-Way Image Transmission Using Corvus Corone Module over WSN. Wireless Pers Commun (2021). https://doi.org/10.1007/s11277-021-08484-2.

[38] Mahajan, H.B., Badarla, A. Cross-Layer Protocol for WSN-Assisted IoT Smart Farming Applications Using Nature Inspired Algorithm. Wireless Pers Commun 121, 3125–3149 (2021). https://doi.org/10.1007/s11277-021-08866-6

[39] Uke, N., Pise, P., Mahajan, H.B., et.al. (2021). Healthcare 4.0 Enabled Lightweight Security Provisions for Medical Data Processing. Turkish Journal of Computer and Mathematics (2021), Vol. 12, No. 11. https://doi.org/10.17762/turcomat.v12i11.5858.

[40] Alhayani, B., Kwekha-Rashid, A.S., Mahajan, H.B. et al. 5G standards for the Industry 4.0 enabled communication systems using artificial intelligence: perspective of smart healthcare system. Appl Nanosci (2022). https://doi.org/10.1007/s13204-021-02152-4.

[41] Mahajan, H.B., Rashid, A.S., Junnarkar, A.A. et al. Integration of Healthcare 4.0 and blockchain into secure cloud-based electronic health records systems. Appl Nanosci (2022). https://doi.org/10.1007/s13204-021-02164-0.

[42] Patil, S., Vaze, V., Agarkar, P. et al. Social context-aware and fuzzy preference temporal graph for personalized B2B marketing campaigns recommendations. Soft Comput (2023).

[43] Kadam, M. V., Mahajan, H. B., Uke, N. J., & Futane, P. R. (2023). Cybersecurity threats mitigation in Internet of Vehicles communication system using reliable clustering and routing. Microprocessors and Microsystems, 102, 104926. https://doi.org/10.1016/j.micpro.2023.104926.