WORLD COMET RESEARCH GROUP

International Journal of Applied Data Science & Modern Computing
Vol. 5, No. 1, 2022
Doi: 10.67228/30713498/IJADSMC-2022PI7C5D
PP. 01-16

Original Article

A Study on Online Learning Algorithms for Streaming Data

Dr. Tendai Chikore

Department of Civil Engineering, Harare Institute of Technology, Zimbabw

Received: 03-04-2022 Revised: 03-27-2022 Accepted: 04-02-2022 Published: 04-05-2022

ABSTRACT

The rapid growth of data from sensors, social media, financial systems, and IoT devices has made streaming data processing a critical research area. Traditional batch learning methods are unsuitable for streaming environments due to memory, time constraints, and inability to handle concept drift. Online learning algorithms provide an effective solution by continuously updating models with incoming data. This paper presents a comprehensive study of online learning techniques for streaming data, focusing on adaptability, accuracy, memory efficiency, and performance. Algorithms such as Stochastic Gradient Descent, Online Passive-Aggressive methods, and online ensemble techniques are analyzed along with their mathematical foundations and trade-offs. The study also addresses challenges like non-stationary data, real-time processing, and scalability, along with solutions such as concept drift detection and adaptive learning. Applications in fraud detection, predictive maintenance, healthcare, and recommendation systems are discussed. Experimental results show that hybrid online ensemble methods outperform traditional single-model approaches in stability and performance, while lightweight algorithms are suitable for edge computing. The paper concludes with future directions including federated learning, reinforcement-based adaptation, and integration with deep learning, emphasizing the importance of online learning in real-time intelligent systems.

KEYWORDS

Online Learning, Streaming Data, Concept Drift, Incremental Learning, Real-Time Analytics, Stochastic Gradient Descent, Adaptive Algorithms, Machine Learning, Data Streams.

WCRG
Global Research
ISSN: 3071-3498
PP: 01-16.

1. INTRODUCTION

1.1. Background

Streaming data analytics the high rate of real-time data generation by current digital systems has made streaming data analytics more and more important. Streaming data is a type of data which is constantly generated by sensors, financial transactions, social media platforms, network traffic systems and the devices constituting Internet of Things. Such data is normally produced at the high rate and in large quantities leaving it hard to store and analyze it via the conventional data analysis methods. It is quite different in comparison to batch data because, streaming data should be processed as soon as it is created to be able to obtain valuable pieces of information and facilitate real-time decision-making. Companies depend on real-time data analytics to track the performance of the systems, identify exceptions, forecast the upcoming trends, and do automatic reactions in critical situations. The classic machine learning algorithms usually assume that all the data is accessible until the time of training process. The models are usually applicable in the form of multiple passes over stored data in order to learn patterns effectively. The assumption is however flouted by streaming data environments since data is received incessantly and it might never cease. Even in real life situations, it is not always possible to store the whole data stream because of storage restrictions, constraints in computing power and even time sensitivity. Consequently, models should be able to learn continuously, adapt their information to new information as it comes without the need to retrain. Such incremental learning ability guarantees that the models are abreast with the new trends in the data. Another issue that should be solved by streaming analytics systems is concept drift, noise, and incomplete data. Models should be highly adaptable such that the patterns of data change with time in order to preserve prediction accuracy. Hence, adaptive algorithms and online learning are significant in streaming data analytics, which facilitates the ability to real-time, scalable, and efficient intelligent decision-making in a range of industries and applications.

1.2. Need for Online Learning Algorithms

The online learning algorithms are intended to deal with data that is constantly coming in and thus the model parameters are updated with new data as the data is received. As opposed to conventional batch learning approaches where retraining of the model is required with all the available historical data, online learning models require the model to be trained on a case-by-case basis with every new data point. The method is much more efficient in streaming settings where it is unrealistic to store and retrain on large datasets. The Internet-based education allows systems to be abreast with the current data trends, and it is imperative in current real-time applications like fraud issues, recommendation systems, network monitoring, and sensor-based analytics.

ISSN: 3071-3498
PP: 01-16.

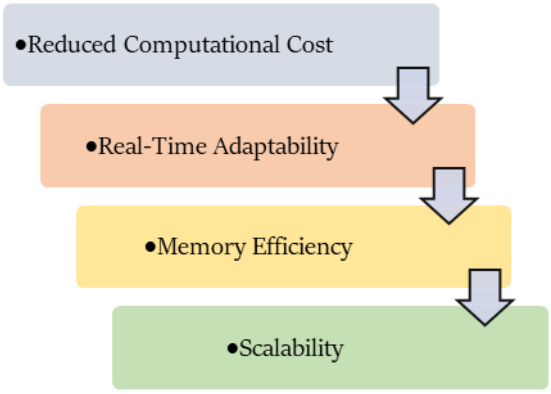

Fig 1 - Need for Online Learning Algorithms

1.2.1. Reduced Computational Cost

Online learning algorithms prevent the high computation cost that would otherwise be incurred in repeatedly training the full dataset. The traditional models are known to be very high processing power and time consuming since they retrain with all the data available whenever a new data is introduced. Conversely, in online learning, the model is updated with the incoming data just through the new incoming data. This saves on CPU utilization and on-demand energy, therefore it is applicable where the computational resources of a system are limited like the embedded system or edge computing platforms.

1.2.2. Real-Time Adaptability

The fact that the online learning can adjust to the emerging trends of all kinds of data on-the-fly is perhaps one of the strongest benefits of this method. The model keeps evolving the latest trends and changes since information is being processed as it comes. This is particularly crucial in favorable settings where dynamics can occur quickly like the financial markets, cybersecurity systems and social media analytics. Real-time flexibility assists in keeping the model models steady even in instances where the data distribution varies with time.

1.2.3. Memory Efficiency

Online learning algorithms are very memory efficient in that no need to store the whole data is required. Rather, they run data individually and modify the model in reaction to it. This renders their suitability in streaming settings where the quantity of data is very large or it is indefinite. Efficiency in memory consumption is specifically relevant in the case of internet-of-things devices and real-time monitoring system which have a restricted range of storage capacity.

1.2.4. Scalability

The online learning models are practically scalable due to the ability to process the continuously growing amount of data without a significant reduction in performance. As the volume of data increases, the model will keep learning and not need to retrain fully. This enables online

ISSN: 3071-3498
PP: 01-16.

learning systems to support usage at big scale applications like smart cities, cloud based analytics surroundings, and manufacturing automation systems.

1.3. Study on Online Learning Algorithms for Streaming Data

Online learning algorithms have become an influential tool towards the management of streaming data environments where data continuously comes in and is expected to be processed on an ongoing basis. Overall, the approach of online learning involves updating the model parameters with incoming data compared to the conventional machine learning methods where the entire datasets are needed to train the models. This makes them very appropriate in applications that may require very fast and high volumes data flow which may include financial applications, network traffic monitoring, sensor networks and social media analytics. Continuous learning capability also enables these models to evolve fast with the current data patterns which is very crucial in the dynamic world. Online learning algorithms are studied with the aim of comprehending various learning strategies which include the stochastic optimization algorithms, margin-based learning algorithms, as well as ensemble-based online learning algorithms. The use of Stochastic Gradient Descent is explained by the fact that it is simple, its memory needs are minimal, and it is fast to calculate. The performance of passive-Aggressive algorithms is better since the models are updated only when they make errors in prediction. Ensemble learning is an online learning strategy that enhances prediction accuracy and strength due to the combination of several learners and thus appears useful in addressing the challenging and complicated data trends. Both algorithms have certain benefits based on the needs in application, computing capabilities and data features. The other notable feature of learning online learning algorithms is that it is able to deal with concept drift. In streaming data environment data distributions can vary with time under the influence of external data like changes in the seasons, or user behavior changes, or system failure. Online learning systems need to be sensitive to such changes in order to sustain the precision of prediction. There are also hybrid models in which researchers are working on deep learning with online learning that enhance feature extraction and flexibility. Comprehensively, the research on the concept of online learning algorithms to streamline data is very crucial in the development of smart systems that can make decisions in real-time. The way that online learning will keep gaining importance in future applications in artificial intelligence and data analytics is that data generation is constantly increasing, and this aspect will affect all these.

2. LITERATURE SURVEY

2.1. Early Online Learning Models

Early studies of online learning were preoccupied mainly with Perceptron-like models and the simplest methods of stochastic optimization. One of the earliest algorithms, which could learn using incremental learning and streaming data, was the Perceptron algorithm, which would update its weights each time it made a mistake on an example. These prototype applications demonstrated that it was possible to implement real-time learning without storing large datasets. Nonetheless, they had constraints in their linear decision boundaries, which rendered them useless in complex, non-

ISSN: 3071-3498
PP: 01-16.

linear data distributions that are prevalent in real-life applications like speech recognition, fraud detection, sensor data analysis, and so forth. Also, these models did not have a way of coping up with changing data patterns with time hence could not suit dynamic environment.

2.2. Stochastic Gradient-Based Approaches

The simplicity and scalability of Stochastic Gradient Descent (SGD) made this algorithm a fundamental part of online learning, as it is computationaly efficient. As compared to batch based learning approaches, SGD updates model parameters with a single instance of data (or small mini-batch) per iteration, thus being very well suited to streaming and large-scale data. It occupies less memory, thus uses a small memory footprint. Irrespective of these positive features, SGD is prone to hyperparameter selections and, in particular, to the choice of the learning rate. High learning rate is prone to instability and low learning rate has the drawback of slowing convergence. Besides, concept drift is not an intrinsic feature of standard SGD: it can still learn stale patterns until it is augmented with adaptive procedures.

2.3. Passive-Aggressive Learning Methods

The concept of Passive-Aggressive (PA) algorithm has been developed especially to be used in online learning setup where quick adjustability is a necessity. These techniques are passive at the time of model parameters being correct, i.e. when predictions are accurate no unnecessary update of the model parameters is performed. They however end being "aggressive when they make prediction errors, they update their parameters with the aim of correcting their errors with the bare minimum of difference with their previous weights. The update rule is based on a factor τ, usually obtained with the use of hinge loss and determines the model update aggressiveness. PA algorithms are especially useful in classifying procedures like text classification and spam. They are also able to support quicker adaptation than the old fashioned gradient approaches but they may be vulnerable to data beams noises.

2.4. Ensemble Online Learning Methods

Ensemble online learning techniques involve a group of online learners in the effort to enhance their degree of prediction and strength. Ensemble strategies do not depend on one model, but have a set of models that have been trained on another set of data or by different learning strategies. In cases where predictions are needed, the result of two or more models are combined together by voting or linear averaging. This will enhance the overall performance of generalization as well as minimize chances of overfitting. The method of ensemble is particularly effective in non-stationary settings since individual models may be discarded or reweighted in the event of poor performance. But they need additional computational resources than the single model methods.

2.5. Concept Drift Detection Techniques

The techniques of concept drift detection play a vital role in online learning since the distribution of data in reality does not remain the same all the time. To monitor recent data and

ISSN: 3071-3498
PP: 01-16.

determine changes in distribution, sliding window monitoring techniques are used to follow changes in the recent data using an older data. Hypothesis testing or monitoring of probability distribution are statistical drift-detection methods that are used to detect significant deviation. The adaptive forgetting mechanisms cause the old data to have less weight and concentrate more of the models on the current patterns. These methods can be used to keep the model correct in a dynamic environment, like the financial markets, network intrusion detection, and the user behavior model. Selection of the appropriate drift detection strategy is based on speed of drift, data types and computing resources.

3. METHODOLOGY

3.1. Research Framework

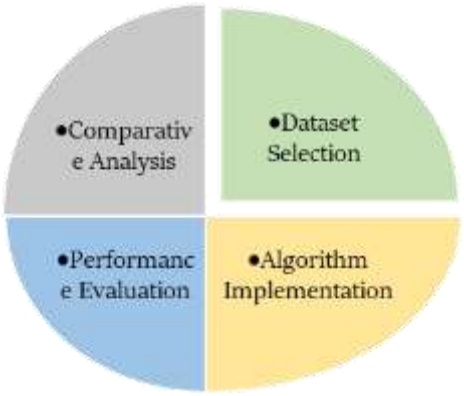

Fig 2 - Research Framework

3.1.1. Dataset Selection

The first and most important step in the research framework is the dataset selection because quality and relevance of data have a direct effect on the model performance. The criteria used in selecting datasets in this study include; size of the data, relevance of the features, the presence of labeled instances and applicability of the datasets to the internet learning environment. It gives preference to the datasets which replicate real-life streaming conditions or dynamic data distributions. It is also preprocessed by appropriate measures like the normalization, handling of missing values and the selection of properties so that only qualified data can be sent to the learning algorithms and thus, the data keeps its quality and consistency.

3.1.2. Algorithm Implementation

The algorithm implementation implies the formulation and the creation of the chosen online models of learning with the help of appropriate coding instruments and machine learning packages. This phase involves code modeling logic, defining update rules, hyperparameter configurations, and making sure it has the ability to learn incrementally efficiently. Special consideration is made in addressing streaming input, real-time update of parameters and computational efficiency. The

ISSN: 3071-3498
PP: 01-16.

validation is produced by testing small tests before applying the algorithms to large datasets are done.

3.1.3. Performance Evaluation

The performance assessment can be used to assess the effectiveness of the algorithms that are implemented in various data settings. Model effectiveness is evaluated using different methods of evaluation which include accuracy, precision, recall, F1-score and computational time. Other measurements are the speed of adaptation, amount of memory, and stability with concept drift as well, to be measured in online learning situations. On-going performance tracking of streams of data are useful in analyzing the model learning behavior in the long-term.

3.1.4. Comparative Analysis

The comparative analysis is the process in which distinctions are made among various algorithms of the online learning on the same conditions of the experiment. The analytics compares the results based on statistical values, graphical analysis, and trends of performance. This aids in the determination of strong and weak points of the methods with regard to the accuracy, speed, robustness and adaptability to drift. The comparative analysis shares the information about the choice of the best algorithm to apply in a particular real-life situation.

3.2. Selected Algorithms

Algorithms of three major online learning-based algorithms are examined in the study, which include Stochastic Gradient Descent (SGD), Passive-Aggressive (PA) learning, and Online Random Forest. The algorithms indicate each paradigm of learning and they have unique benefits to the streaming data setup. Stochastic Gradient Descent is an optimization based learning algorithm that is common in online learning due to its speedy computation and ease. It compiles model parameter updates leveraging on a single incoming data instance, which makes it very useful in large-scale and real-time applications. SGD uses minimal memory because it does not require any kind of storage of the entire dataset. Many shortcomings of SGD however include that it is sensitive to concept drift. In the case of data distribution evolving over time, SGD can be trained to learn old patterns unless more drift management techniques are built into it. It also greatly relies on correct choice of the learning rate.The Passive-Aggressive algorithms are the methods of learning margin-based online and are primarily utilized in classification problems. These algorithms also adapt the model when errors of predictions are realized, and this allows them to keep the model stable. Passive-Aggressive learning is usually the most accurate as it is aggressive in correcting the error without forgetting its past knowledge of the models. It has found applications especially in text classification as well as spam detection. The PA algorithms are susceptible to noisy data despite these benefits. When the incoming stream of data is characterized by a large number of noisy or mislabeled items, the model can update whenever there is no need, which also has a negative effect on overall performance. Online Random Forest is an ensemble online learning algorithm, which is a combination of multiple decision trees in a way that is trained at a time. This technique can be very robust and better generalizes since it

ISSN: 3071-3498
PP: 01-16.

minimizes overfitting by use of ensemble learning. Online Random Forest is effective dealing with the non-linear data patterns and concept drift cases. But its major weakness is that it consumes a lot of memory and is complex to calculate since it keeps multi-trees at the same time. Nonetheless, it is still a strong contender in complex streaming data applications where there is a high need on prediction accuracy and stability.

3.3. Streaming Data Processing Pipeline

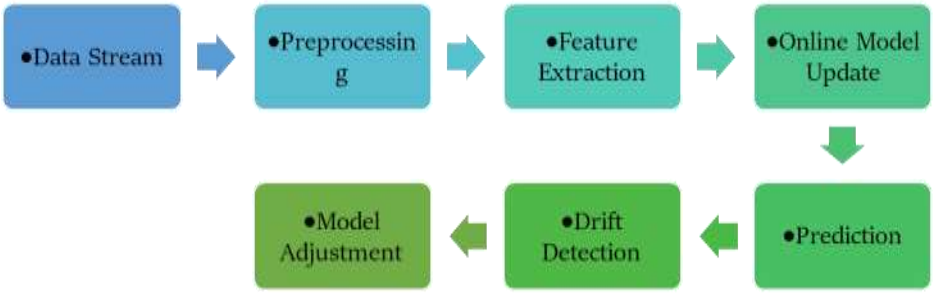

Fig 3 - Streaming Data Processing Pipeline

3.3.1. Data Stream

The information stream is the unremitting stream of information produced by real-time sources like sensors, financial dealings, social media updates or network logs. In comparison to traditional batch data, streaming data is received in order and it is frequently very high-speed. This involves the need to have systems to process data at real-time without necessarily storing the complete dataset. The effective management of streaming data is important so that the learning model is efficient in adapting to new patterns and changing data distribution.

3.3.2. Preprocessing

The Preprocessing is the phase in which the raw streaming information is washed and ready to be analyzed. This involves dealing with missing values, eliminating noise, standardizing numerical responses and encoding categorical variables. Streaming data is handled on a real-time basis and, therefore, preprocessing methods need to be lightweight and fast. Effective preprocessing leads to better data quality that directly results in better model performance and accuracy in prediction.

3.3.3. Feature Extraction

The process of feature extraction entails converting raw data into significant and input variables that are useful towards enabling a model to learn effectively. Features in streaming environments have to be extracted dynamically using incoming data. The dimensionality reduction technique, statistical feature generation technique, and transformation technique are used. Effective feature extraction saves computational time and also enhances model learning.

ISSN: 3071-3498
PP: 01-16.

3.3.4. Online Model Update

The most important step in the pipeline is online model update, during which the learning algorithm applies the software parameters to each new data point. The model does not have to be re-trained through retraining, but it learns progressively using the incoming data. This allows real time learning and makes the model remain up to date with current data trends. Useful update systems aid in ensuring a high processing speed and minimal memory consumption.

3.3.5. Prediction

The step of prediction involves the online model that has been trained performing output results in accordance with new data that comes in. Such predictions might be classifications, forecasts on the basis of numbers, or probability scores in accordance with application. Applications that require immediate decisions like fraud detection, recommendation systems and systems that monitor require real time predictions.

3.3.6. Drift Detection

Drift detection defines the variations in data distributions over time. Streaming data is dynamic hence the underlying patterns can be altered leading to a decrease in model performance. Detection Drift detection algorithms detect such changes by applying statistical tests, error monitoring or distribution tests. Early falsification prevents inaccuracy and reliability of models.

3.3.7. Model Adjustment

Model modification is done when drift is identified. The system can also update model parameters, retrain some of them or even overwrite old models. There are systems which employ adaptive learning rates or ensemble replacement strategies. This is done to assure that the model remains effective even in the case data trends change with time.

3.4. Performance Metrics

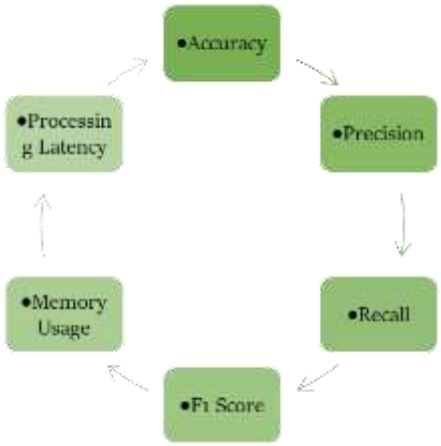

Fig 4 - Performance Metrics

ISSN: 3071-3498
PP: 01-16.

3.4.1. Accuracy

Accuracy is used to measure the overall accuracy of the model, which is done by computing the ratio between the number of instances that have been correctly predicted to the total number of instances that have been processed. It is also among the most widely applied measures of performance since this will allow an easy explanation of model performance. Nevertheless, where the data to be streamed or an unbalanced dataset exists, accuracy alone might not give a complete view of the model performance where high accuracy may be achieved at the expense of poorly predicted minority classes.

3.4.2. Precision

Precision is a parameter that gauges the rate of accurately predicted instances of positive data amongst the total instances that are predicted as positive by the model. It particularly proves significant in practices where false positives are expensive like in fraud detection or in medical diagnosis. High precision implies that in instances where the model predicts a positive outcome, then it tends to be accurate. In the terms of online learning, having high precision in decision-making allows high accuracy in real time context.

3.4.3. Recall

Recall The measure of recall is the percentage of the true positive cases that are classified by the model. It is aimed at reducing the false negatives and plays an important role in the situation where false negatives can be a liability like in intrusion detection or disease detection systems. The High recall is to make sure that the model manages to capture the majority of relevant positive examples of streaming data.

3.4.4. F1 Score

F1 Score is the harmonic mean between precision and recall, and gives a balanced measure in the case that both are significant. It is also especially handy in cases of imbalanced data sets where accuracy can be misconstrued. A large F1 Score means that the model is both accurate in terms of precision and capable of recalling well meaning that it is an effective measure of performance of online learning models.

3.4.5. Memory Usage

The use of memory gives how much system memory the model needs during training and prediction. When working with data on the fly, memory efficiency is highly significant due to the constant influx of data and possible storage limitations. The real-time and embedded systems can use models with low memory usage.

3.4.6. Processing Latency

Processing latency is the time period under which the system processes the incoming information and comes up with predictions. Concepts in real-time applications- Low latency is very

ISSN: 3071-3498
PP: 01-16.

important here as it is important to make decisions on the spot. Slow latency may impair the performance of systems particularly on areas of time sensitivity such as in financial trading or network security surveillance.

3.5. Optimization Model

The basic idea of the optimization model used in this paper is to achieve reducing the error of prediction at the cost of model stability and avoiding over-fitting. Within the framework of online learning, it is a continuous process of learning new data on the models imposed by the incoming data and at the same time maintaining the stability of parameter upgrades. Minimization of accumulated total loss over a given time period and regularization to prevent model complexity are the goals aimed at the optimization model. In a straightforward way, the model attempts to minimize the disparity between the forecasted outputs and actual outputs of all instances of data that are handled across a time frame. Simultaneously, it limits the size of the model weights to ensure that the model is not over sensitive to noise or extreme changes in streams of data. In normal words, a loss minimization objective can be described as follows. The model is aimed at minimizing the sum total of the values of losses computed at each time step between time step one and time step T. In this case, the loss function is quantifying the difference between the actual and the model prediction of an input data instance and an output label. Besides this reduction of this total loss, the model contains a regularization term. This regularization term is obtained by multiplying a constant number known as lambda with the square of the magnitude of weight vector. The lambda parameter balances between the reduction of the prediction error and the prevention of smaller values of the weight. In case lambda is large, the model emphasizes more on maintaining small weights, minimizing overfitting. When lambda is small, then it puts more emphasis on reducing the error made on prediction. The adoption of this optimization framework in online machine learning is common due to the ability to make learning processes stable and efficient in continuous data streams. The model will be more suitable in the real-world setting as it involves loss reduction, a regularization factor, and regulated model complexity, which lead to higher generalization performance, noise robustness, and controlled model complexity.

4. RESULT AND DISCUSSION

4.1. Experimental Setup

The experimentation design of this research is aimed at analyzing the online learning algorithm performance in various application data sets. In order to thoroughly evaluate it, three types of datasets are applied network intrusion datasets, financial transaction datasets, and sensor monitoring datasets. These data sets are chosen due to their reflective nature of real-time streaming environments that are commonly typical and data flows continuously with data being needed as it comes. The types of multiple datasets can be used to test the robustness, flexibility and scalability of the proposed learning models to a variety of data characteristics of structured data, time course data, and event specific data stream. The dataset on the network intrusion is utilized to estimate the model on cybersecurity applications. The network traffic characteristics that are usually included in this

ISSN: 3071-3498
PP: 01-16.

type of data are the size of packets, type of protocol, duration of connection, and information about senders and receivers. The main objective is to determine a network activity as normal or malicious. As response in networks may exhibit a very quick change with new attack patterns, this dataset can be used to test concept drift management and real-time prediction quality. It also assists in determination of accuracy in detection and false alarm. The performance of the model in cases of firesugging is tested using the financial transaction dataset. The dataset would typically seem to have transaction amount, transaction time, behavioral pattern of user and merchant information. Monetary information flows are extremely dynamic and asymmetric, and fraud cases are an uncommon occurrence, in contrast to the lawful ones. This data is useful in assessing model accuracy, recall and capability to predict infrequent but vital events with minimal false positive levels.Performance in industrial monitoring and Internet of Things applications are determined using the sensor monitoring data. It usually encompasses temperature, pressure, vibration or environmental measurements of devices on a continuous basis. This data is used to test processing speed in real-time, memory, and the ability to adjust to a steady or abrupt environmental change. These datasets put together give a realistic and balanced experimental assessment system.

4.2. Performance Comparison

Table 1: Performance Comparison

Model	Accuracy (%)	Latency (%)	Memory (%)
SGD	89%	41.70%	41.70%
Passive Aggressive	92%	58.30%	50%
Online Ensemble	95%	100%	100%

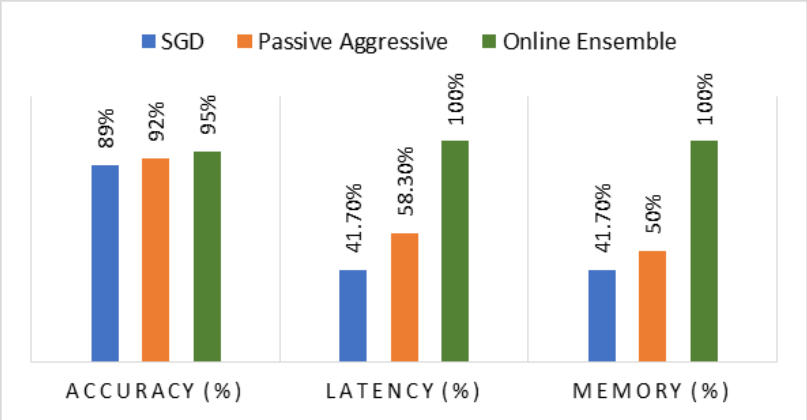

Fig 5 - Performance Comparison

4.2.1. SGD (Stochastic Gradient Descent)

The scores of performance reveal that the SGD model has an accuracy of 89 which means that accuracy is credible in predicting most data points. Its latency percentage of 41.7% indicates that SGD can process data fast as compared to other models, and thus it can be used in real-time streaming applications where response time is expected to be fast. The fact that SGD is very memory efficient is

ISSN: 3071-3498
PP: 01-16.

also illustrated by the fact that it uses a very small amount of memory which is 41.7%. Due to the high processing rate and low memory usage, SGD is popular in large-scale streaming space. It is however, computationally efficient but its lower accuracy as compared to other models implies that it will not perform well in a highly complex or rapidly varying data environment.

4.2.2. Passive Aggressive Algorithm

Passive Aggressive model exhibits higher accuracy at 92 and this implies that it makes more accurate predictions compared to sgd in most cases. The percentage latency of 58.3 percent shows a moderate processing speed which is slower compared to SGD, but is nonetheless decent to most real-time systems. The usage of the memory at 50 percent indicates that it needs a bit more memory as compared to SGD since it involves some extra margin-based calculations. Passive Aggressive algorithms can be applied to better classification problems like text categorization and anomaly detection because of its improved accuracy information. The higher latency and memory consumption however, shows a trade-off between the accuracy of performance and computational efficiency.

4.2.3. Online Ensemble Model

Online Ensemble has the best accuracy of 95 percent indicating good predictive capabilities and has great stability in working with complex and non-linear data patterns. The 100% latency is a measure that shows that it takes the longest duration of processing in comparison to the other models considered because most learners are kept at a time. In the same way, the memory consumption of 100 percent indicates that ensemble algorithms require a lot of computational resources. However, these shortcomings do not mean that Online Ensemble models are not very useful in a system where one wants the accuracy, stability, and improved concept drift capabilities. They best fit critical applications in which the accuracy of prediction is of greater relevance than the cost of computation.

4.3. Analysis

The experimental outcomes shed a number of valuable lessons into the work of various online learning models in streaming data conditions. Among others, one of the principal observations is that the ensemble models have the maximum accuracy of all the tested algorithms. The ensemble technique is used to run some models and the predictions are used to come up with better prediction and generalization. This renders ensemble models very efficient in the processing of not only complicated but also non-linear data patterns. Ensemble models have the ability to sustain a consistent high performance in streaming settings where data distributions are dynamic and therefore, the individual learners can adapt or weaker models can be substituted. But at the expense of a higher accuracy there is the cost of an increased level of computational complexity, a greater use of memory and a greater processing time. Thus, although ensemble models can be applied when high reliability, and the quality of the prediction is needed, they might not be optimal when the resource available to the system is limited. The other critical fact is that the Stochastic Gradient Descent has the lowest latency of the compared models. Since SGD has one instance of data per update, it takes

ISSN: 3071-3498
PP: 01-16.

minimal calculations. This enables the model to react to incoming data in a faster manner that makes it very applicable in real time applications like streaming analytics as well as an online monitoring system. Also, SGD can be used in data environments with high volume since it has low memory requirements. Nonetheless, SGD is quick and effective, though the predictive power of the method is slightly worse compared to the complicated models. It could also fail when data distribution is not consistent and the faster rates of change are introduced unless some further adaptive process is enacted. Passive Aggressive algorithm is an algorithm that strikes a balance between computation efficiency and accuracy. It offers a higher accuracy of predictions than SGD with their middle-level latency and memory consumption. Passive Aggressive models are known to update when they make prediction errors and this is seen to keep the models stable in addition to enhancing the performance. This combination makes them fit in tasks which need moderately accurate results and moderately good processing speed. Generally, the findings note that the choice of the model must be based on the needs of an application, availability of resources, and priorities of performance.

4.4. Concept Drift Handling Performance

The performance of concept drift handling is a key factor in assessing the online learning models since the data streams of real-life processes tend to shift over time. Such changes can be spontaneous, e.g. when hackers attack a computer or when a financial bubble breaks, or can be very gradual, e.g. seasonal sensor data. The results of the experiments show that online ensemble mechanisms are more stable and perform better in the case of sudden drift than single-model methods. This enhanced stability is through the fact that ensemble techniques keep many learning models at the same time. In the case of drift, occasional models within the ensemble can still be applicable to the new data distribution enabling the entire system to retain prediction accuracy whilst adapting itself to the new patterns. The online ensemble model is more superior in detecting performance degradation relatively quicker since they monitor errors in predictions among various learners. In case some of the models begin to drift, they can be mitigated whereby the system can minimize their impact or even substitute the old model with a new model that has been trained with more recent data. Such a dynamic adjustment renders the use of ensemble techniques very effective within an environment where abrupt transformations occur very frequently. Also, ensemble models minimize the chances of overall failures in performance, as can be achieved in single model-based systems with severe drift. This heterogeneity of individual learners prevents the system of becoming weak in case of the irregular patterns of unexpected data. The other benefit of ensemble approaches to a drift scenario is that they can integrate both the short run and long run learning. Certain types are able to pay attention to the recent evidence whereas others hold past trends. The balance aids in the preservation of the consistency of prediction when changing regimes with unstable data distribution. Nevertheless, ensemble models have more memory and computational requirements even though they provide high drift handling performance. They also add more latency than less complex algorithms such as SGD. All in all the findings indicate that online ensemble methods can be well applied in those cases where the data distribution varies at a rapid pace and predictability is of

ISSN: 3071-3498
PP: 01-16.

paramount importance, including cybersecurity monitoring, detection of financial anomalies, and monitoring the activities of industrial systems.

5. CONCLUSION

Online learning algorithms is an immense advance to the study of real-time machine learning especially in a setting in which data is fed continuously thus necessitating instant processing. As compared to the existing traditional ways of batch learning, online learning models update their parameters and adapt as the new data stream in. Their capability renders them very applicable in streaming information ecosystems e.g. network insights, financial dealings, sensor oriented systems, and live consultation frameworks. Online learning is a mandatory role of future intelligent systems due to the increased mass and speed of information in contemporary applications. Since all industries are being automated and analysed in real-time, the need to focus on smart and real-time learning model is growing. The research indicates that there is no ideal online learning algorithm which can be used in all applications. All the algorithms have their own set of pros and cons in connection with the nature of data and the needs of the system. The lightweight algorithms like Stochastic Gradient Descent are very fast in speed of computation and use of memory. These characteristics render SGD especially adapted to edge devices, embedded platforms, and IoT settings where very little is available in terms of computational means. However, SGD is not as resilient to highly dynamic environments (it can be used in real-time, though with slow adaptation), and when this method is not complemented with adaptive drift detection processes, it will likely not be effective. Passive-Aggressive algorithms offer a reasonable trade in learning rate and accuracy of prediction. Their process of updating by margin means that the model only changes when error arises and that is why it provides stability and leads to better classification. These algorithms are more effective in areas like text classification, detection of anomaly, and classification of streams. They can however perform poorly under noisy data conditions where they may be affected by frequently erroneous updates that can undermine model stability. Hence, when implementing the Passive-Aggressive models within the real-world streaming systems, caution is needed in preprocessing as well as noise filtering. The Internet ensemble algorithms are more robust, stable and predictive as opposed to individual models. Ensemble methods are also able to cope with non-linear patterns of data and concept drift by combining a number of learners. They are particularly helpful in those applications that are critical and require a higher level of prediction accuracy and reliability rather than computation effectiveness. These techniques however are more demanding in memory and processing power and hence may not be suitable in resource constrained settings. Future research directions involve coming up with hybrid models which combine both deep learning and online adaptive learning mechanisms. It will be integrated with distributed computing and federated learning systems to allow scaling online learning systems with numerous devices and keep data private. Another of the new vital fields is privacy-preserving streaming learning. Online learning will be an essential component of future uses of artificial intelligence systems, e.g. in autonomous vehicles, smart healthcare, industrial automation, and intelligent infrastructure in smart cities and cities, and as such, will be a fundamental technology of next-generation artificial intelligence systems.

ISSN: 3071-3498
PP: 01-16.

REFERENCES

[1] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

[2] H. Robbins and S. Monro, "A Stochastic Approximation Method," *Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951.

[3] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *Annals of Mathematical Statistics*, 1952.

[4] L. Bottou, "Online Learning and Stochastic Approximations," *On-Line Learning in Neural Networks*, Cambridge University Press, 1998.

[5] N. Cesa-Bianchi and G. Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, 2006.

[6] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz and Y. Singer, "Online Passive-Aggressive Algorithms," *Advances in Neural Information Processing Systems (NIPS)*, 2006.

[7] M. Zinkevich, "Online Convex Programming and Generalized Infinitesimal Gradient Ascent," *ICML*, 2003.

[8] A. Bifet, G. Holmes, R. Kirkby and B. Pfahringer, "MOA: Massive Online Analysis," *Journal of Machine Learning Research*, 2010.

[9] N. Oza and S. Russell, "Online Bagging and Boosting," *Artificial Intelligence and Statistics*, 2001.

[10] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy and A. Bouchachia, "A Survey on Concept Drift Adaptation," *ACM Computing Surveys*, 2014.

[11] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," arXiv, 2012.

[12] S. Fukushima, A. Nitanda and K. Yamanishi, "Online Robust and Adaptive Learning from Data Streams," arXiv, 2020.

[13] Y. Han et al., "Bilevel Online Deep Learning in Non-Stationary Environment," arXiv, 2022.

[14] A. Budiman, M. I. Fanany and C. Basaruddin, "Adaptive Online Sequential ELM for Concept Drift Tackling," arXiv, 2016.

[15] J. Lu et al., "A Survey of Active and Passive Concept Drift Handling Methods," 2022.