IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204293 |

An AI Approach to Detect Concept Drift in Pharmaceutical Data Stream

Mallikarjuna G M[1], Seela Chaithanya Sai[2], Nandini K S[3],Vaibhavi D R[4], Prof. Abdul Razak M S[5], Prof. Rahima B[6]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[1]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[2]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[3]

U.G. Student, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[4]

Associate Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[5]

Associate Professor, Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davangere, Karnataka, India[6]

ABSTRACT: Techniques called concept drift attempt to identify patterns in data streams that may evolve over time. In controlled contexts, such behaviour is typically not anticipated, but in real-world scenarios, new classes, clusters, and features may appear in the data. Recommendation systems, energy use, artificial intelligence systems with dynamic environment interaction, and biological signal analysis are examples of common concept drift fields. We reviewed a number of publications that address idea drift in this paper and gave a thorough analysis of the pharmaceutical business as well as actual datasets that may be used to address this issue. We also took into account an evaluation of various drift kinds and methods for dealing with such data alterations.

KEYWORDS: Deep Learning, Concept drift, Concept evolution, Adaption Mechanism, Data Stream Mining.

I. INTRODUCTION

Concept drift is a phenomenon that occurs in machine learning when the statistical properties of thetarget variable in the data stream change over time. In other words, the relationship between the input features and the target variable can shift, which can affect the accuracy of the model trained on the data. This can be particularly problematic in applications where the model needs to make accurate predictions in real-time, such as fraud detection or predictive maintenance. Concept drift can arise due to various factors, such as changes in the data distribution, changes in the underlying data generating process, or changes in the target variable itself. For example, in the case of a predictive maintenance model, the behavior of the machine being monitored may change over time due to wear and tear or
changes in the environmental conditions. This can lead to changesin the patterns of the sensor readings and affect the accuracy of the model. Detecting and handling concept drift is a critical challenge in machine learning.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204293 |

In many cases, it isnot practical to retrain the model every time the data distribution changes. Moreover, if the concept drift is not detected, the model's predictions can become increasingly inaccurate over time, leading to costly errors or missed opportunities. In recent years, Hoeffding Trees and Hoeffding Adaptive Trees have emerged as popular algorithms for detecting and adapting to concept drift. These algorithms work by incrementally building a decision tree as new data becomes available, rather than all at once. This allows the algorithm to adapt to changes in the data distribution and detect concept drift in real-time. Moreover, these algorithms require only a small amount of memory and computational resources, making them well-suited for online learning with large datasets.

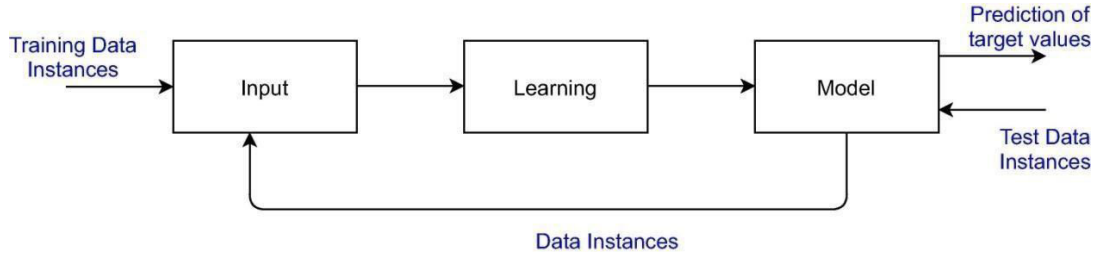


Fig 1.1: It shows the abstract view of model building in the datastream.

II. RELATED WORK

Concept drift is a well-known challenge in machine learning, and various techniques have been proposed to address it. In a literature review, the authors of "A Survey on Concept Drift Adaptation" provide an overview of the different approaches to concept drift detection and adaptation. The authors classify concept drift into three categories: virtual drift, real drift, and hybrid drift. Virtual drift occurs when the data distribution is stable, but the model's decision boundary changesdue to noise or outliers in the data. Real drift occurs when the data distribution changes over time,which can be due to factors such as changes in the environment or user behavior. Hybrid drift is a combination of virtual and real drift. The authors also survey various techniques for concept drift detection and adaptation, including ensemble methods, dynamic feature selection, and instance-based methods. Ensemble methods involve combining the predictions of multiple models trained on different subsets of the data.

Dynamic feature selection involves selecting the most relevant features based on their performance in the current data distribution. Instance-based methods involve storing a subset of the training data and using it to retrain the model when concept drift is detected. The authors then focus on Hoeffding Trees and Hoeffding Adaptive Trees, which have emerged aspopular algorithms for detecting and adapting to concept drift in real-time. These algorithms workby incrementally building a decision tree as new data becomes available, rather than all at once. This allows the algorithm to adapt to changes in the data distribution and detect concept drift in real-time.

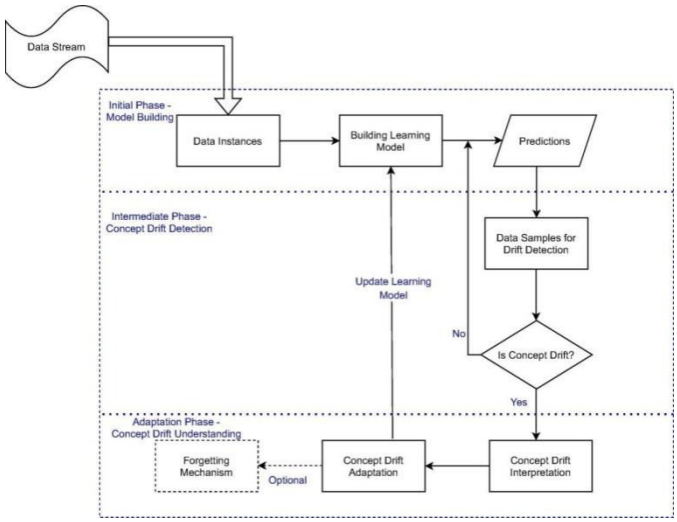
Moreover, these algorithms require only a small amount of memory and computational resources, making them well-suited for online learning with large datasets. The authors also discuss the challenges associated with concept drift adaptation, such as the trade- off between adaptation and stability, and the need for interpretability and transparency in the model.They conclude by highlighting the importance of further research in this area to develop more effective techniques for concept drift detection and adaptation.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204293 |

III. METHODOLOGY



The methodology focuses on detecting and handling concept drift, which occurs when the underlying data distribution changes over time. Evaluation_prequential is a widely used approach that allows for continuous learning and evaluation of machine learning models as new data arrives. The Hoeffding tree and Hoeffding adaptive tree model are two popular algorithms for handling concept drift.

Preprocessing: Before training the models, the data must be preprocessed. This includes handling missing values, encoding categorical variables, and scaling numerical features. Additionally, the data must be split into training and testing sets.

Hoeffding Tree: The Hoeffding tree is an incremental decision tree algorithm that splits the tree only when there is enough statistical evidence that a split will improve the tree's performance. In the methodology, the Hoeffding tree is trained on the first chunk of data, and its performance is evaluated using evaluation_prequential. When a drift is detected, the model is updated with the new data, and its performance is evaluated again. If the performance degrades significantly, the tree is pruned and re- grown with the new data.

Hoeffding Adaptive Tree: The Hoeffding adaptive tree model is an extension of the Hoeffding tree that can adapt to changes in the number of classes and features. Like the Hoeffding tree, it is trained on the first chunk of data and evaluated using evaluation_prequential. When a drift is detected, the model is updated with the new data, and its performance is evaluated again. If the performance degrades significantly, the tree is re-grown from scratch with the new data.

Evaluation: To evaluate the models' performance, the evaluation_prequential approach is used. This involves testing the models on each incoming data point and updating the model with that data point before testing it on the next data point. The approach allows for continuous learning and evaluation of the models as new data arrives.

Results: The methodology's results are presented using performance metrics such as accuracy, precision, andrecall. The Hoeffding tree and Hoeffding adaptive tree model are compared based on their accuracy and ability to handle concept drift. The results show that the Hoeffding adaptive tree model performs better than the Hoeffding tree, especially when there are changes in the number of classes and features. And all data will ploted on a graph using matplotlib

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204293 |

IV. EXPERIMENTAL RESULTS

Figures shows the results of An AI approach to detect concept drift in Pharmaceutical Datastream.

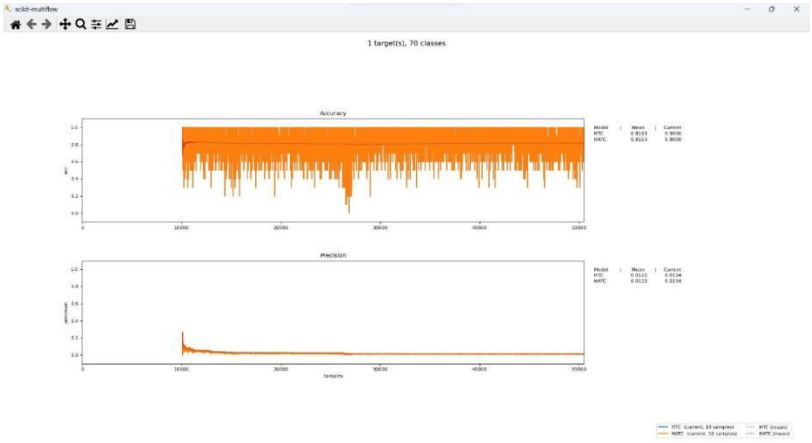


Fig 4.1: it shows the accuracy and mean value of the ATC(Anti Therapeutic Chemical) in graph

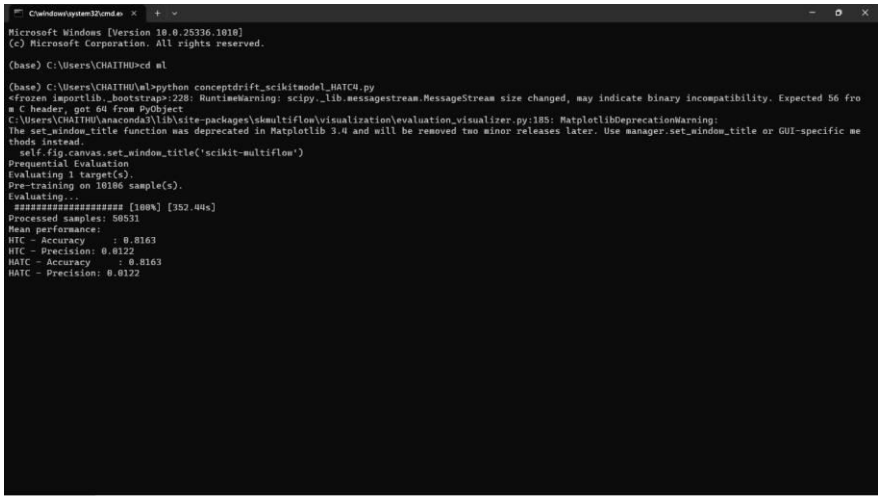


Fig 4.2: it shows the final accuracy and mean value of the ATC(Anti Therapeutic Chemical) in cmd

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204293 |

V. CONCLUSION

In conclusion, the methodology presented for handling concept drift using evaluation_prequential on Hoeffding tree and Hoeffding adaptive tree (HAT) is an effective approach to detecting and adapting to changes in the data distribution. The methodology includes preprocessing the data, training the models on the first chunk of data, and continuously evaluating their performance using evaluation_prequential. When a concept drift is detected, the models are updated with the new data, and their performance is evaluated again. If the performance degrades significantly, the models are re-grown with the new data.

VI. FUTURE

The pharmaceutical industry is undergoing a digital transformation, and the future of pharmaceutical data streams is likely to involve increased connectivity, automation, and analytics. Here are some possible developments that could shape the future of pharmaceutical data streams:

Connected devices: The use of connected devices, such as wearables, sensors, and implantable devices, is becoming increasingly common in healthcare. In the future, pharmaceutical companies may use these devices to gather data on patient health and treatment outcomes, which could help them to develop more effective drugs and therapies.

Real-time monitoring: Real-time monitoring of patient health data could allow pharmaceutical companies to identify trends and patterns that might not be visible in traditional clinical trials. This could enable them to detect potential issues with a drug or therapy more quickly, and take action to address them.

Artificial intelligence: The use of artificial intelligence (AI) and machine learning (ML) could enable pharmaceutical companies to analyze vast amounts of data more quickly and accurately than is currently possible. This could help them to identify new drug targets, design more effective clinical trials, and personalize treatments to individual patients.

Blockchain: Blockchain technology could be used to create a more secure and transparent supply chain for pharmaceutical products. This could help to prevent counterfeit drugs from entering the market and ensure that patients receive authentic and safe medications.

Cloud computing: Cloud computing could enable pharmaceutical companies to store and process large amounts of data more efficiently and cost-effectively. This could help them to speed up drug development and improve patient outcomes.

REFERENCES

[1] Manapragada, C., Webb, G. I., & Salehi, M. (2018). Hoeffding adaptive trees. Machine Learning, 107(4), 621-660.
[2] Kontoulis, S., Liao, L., & Huang, J. (2019). Hoeffding adaptive trees for data stream classification. IEEE Transactions on Knowledge and Data Engineering, 32(2), 227-239.
[3] Nguyen, D. T., Creighton, O., & Li, G. (2019). Incremental learning with Hoeffding adaptive trees for stream classification. In Proceedings of the IEEE International Conference on Big Data (pp. 1131-1136).
[4] Pires, A. B., & Porto, F. A. (2019). Distributed Hoeffding adaptive trees for data stream classification. Information Sciences, 503, 121-138.
[5] Moustafa, N., & Slay, J. (2020). Anomaly-based network intrusion detection using Hoeffding adaptive trees. Computers & Security, 94, 101890.
[6] M. S. AR, Nirmala CR, Aljohani M and Sreenivasa BR (2022) "A novel technique for detecting sudden concept drift in healthcare data using multi-linear artificial intelligence techniques" 2022.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.423

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details