

## Chapter 2

# A Survey of Class Imbalance Problem on Evolving Data Stream

**D. Himaja**

*Vignan's Foundation for Science, Technology, and Research (Deemed), Guntur, India*

**T. Maruthi Padmaja**

*Vardhaman College of Engineering, Hyderabad, India*

**P. Radha Krishna**

 <https://orcid.org/0000-0002-2764-2818>

*National Institute of Technology, Warangal, India*

### ABSTRACT

*Learning from data streams with both online class imbalance and concept drift (OCI-CD) is receiving much attention in today's world. Due to this problem, the performance is affected for the current models that learn from both stationary as well as non-stationary environments. In the case of non-stationary environments, due to the imbalance, it is hard to spot the concept drift using conventional drift detection methods that aim at tracking the change detection based on the learner's performance. There is limited work on the combined problem from imbalanced evolving streams both from stationary and non-stationary environments. Here the data may be evolved with complete labels or with only limited labels. This chapter's main emphasis is to provide different methods for the purpose of resolving the issue of class imbalance in emerging streams, which involves changing and unchanging environments with supervised and availability of limited labels.*

DOI: 10.4018/978-1-7998-7371-6.ch002

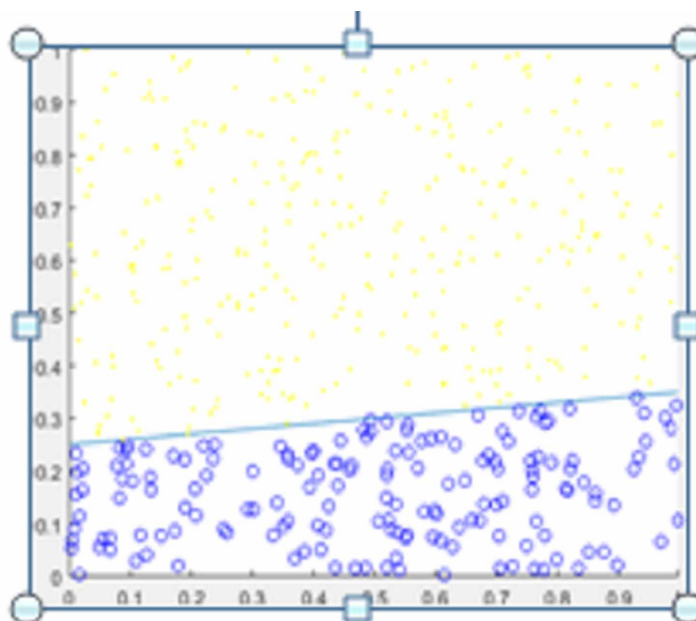
## INTRODUCTION

The real-world classification problems, such as fraud and fault detection are characterized by continuously imbalanced evolving streams from non-stationary environments. The combined problem of class imbalance and concept drift, on the other hand, hinders the success of online learners. Learning from unbalanced emerging streams poses a different set of problems than learning from balanced groups. The problem of learning from emerging streams, on the other hand, faces significant challenges such as infinite stream size, varying speed, and concept drift. The method of extracting information from continuous, rapid data records is known as stream learning. It's a huge challenge to learn from these streams. The evolving data may be completely or partially labeled.

### Class Imbalance Learning

On standalone training sets, the class imbalance learning (CIL) problem arises when one class of data vastly outnumbered the others, causing the underrepresented class output to suffer. Therefore the classes from over represented class are correctly classified where as from under represented are misclassified. This scenario is common in applications such as fraud and fault detection. Smart building is an example for imbalanced data. It has sensors to detect risky conditions. Any fault in sensors causes a great destruction. The size of smart building is 5000 with two classes namely faulty and good, where only 1% is faulty and remaining 99% is good conditions. A model can be built on this kind of datasets, to predict faults in sensors. But predicting faults is expensive as faulty conditions are underrepresented than good. Figure 1 shows the scenario for class imbalance. Here there are two classes (i.e. blue and yellow). From the figure 1, it is clear that the dataset is highly imbalanced where blue is 70% and remaining 30% is yellow class.

Figure 1. Class imbalance



17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-survey-of-class-imbalance-problem-on-evolving-data-stream/280909](http://www.igi-global.com/chapter/a-survey-of-class-imbalance-problem-on-evolving-data-stream/280909)

## Related Content

---

### JIRL: A C++ Toolkit for JPEG Compressed Domain Image Retrieval

David Edmundson and Gerald Schaefer (2013). *International Journal of Multimedia Data Engineering and Management* (pp. 1-12).

[www.irma-international.org/article/jirl/84022](http://www.irma-international.org/article/jirl/84022)

### Exponential Organizations

Presleyson Plínio de Lima (2023). *Enhancing Business Communications and Collaboration Through Data Science Applications* (pp. 107-138).

[www.irma-international.org/chapter/exponential-organizations/320753](http://www.irma-international.org/chapter/exponential-organizations/320753)

### Modeling Joint Visual Attention in Naturalistic Dyadic Interactions

Kumushini Thennakoon, Yasasi Abeysinghe, Bhanuka Mahanama, Vikas Ashok and Sampath Jayarathna (2026). *International Journal of Multimedia Data Engineering and Management* (pp. 1-19).

[www.irma-international.org/article/modeling-joint-visual-attention-in-naturalistic-dyadic-interactions/406715](http://www.irma-international.org/article/modeling-joint-visual-attention-in-naturalistic-dyadic-interactions/406715)

### Blockchain-Secured Electronic Health Records for the Internet of Medical Things

Poonguzhali N. (2024). *Cybersecurity and Data Management Innovations for Revolutionizing Healthcare* (pp. 260-279).

[www.irma-international.org/chapter/blockchain-secured-electronic-health-records-for-the-internet-of-medical-things/351714](http://www.irma-international.org/chapter/blockchain-secured-electronic-health-records-for-the-internet-of-medical-things/351714)

### DMMs-Based Multiple Features Fusion for Human Action Recognition

Mohammad Farhad Bulbul, Yunsheng Jiang and Jinwen Ma (2015). *International Journal of Multimedia Data Engineering and Management* (pp. 23-39).

[www.irma-international.org/article/dmms-based-multiple-features-fusion-for-human-action-recognition/135515](http://www.irma-international.org/article/dmms-based-multiple-features-fusion-for-human-action-recognition/135515)