# Chapter 2
# Knowledge Discovery From Evolving Data Streams

**Prasanna Lakshmi Kompalli**
*Gokaraju Rangaraju Institure of Engineering and Technology, India*

## ABSTRACT

*Data coming from different sources is referred to as data streams. Data stream mining is an online learning technique where each data point must be processed as the data arrives and discarded as the processing is completed. Progress of technologies has resulted in the monitoring these data streams in real time. Data streams has created many new challenges to the researchers in real time. The main features of this type of data are they are fast flowing, large amounts of data which are continuous and growing in nature, and characteristics of data might change in course of time which is termed as concept drift. This chapter addresses the problems in mining data streams with concept drift. Due to which, isolating the correct literature would be a grueling task for researchers and practitioners. This chapter tries to provide a solution as it would be an amalgamation of all techniques used for data stream mining with concept drift.*

## INTRODUCTION

Innovations in digital era has given rise to huge amounts of data. Every day activities like banking transactions, web browsing, mobile usage, communication networks and even hardware devices generate huge amount of flowing data. The amount of data transferred in varieties of channels is growing exponentially. Streams of data so generated contains knowledgeable information. This information can be used in decision making, development of better quality products, finding new relations

among existing items etc. As it is not conceivable to manually study all the data, automated techniques must be developed with good computational power for finding valuable and relevant information. More over analyzing data which is changing or evolving over a period time while flowing is a challenge for real time analytics. Data Streams which evolve or change are termed as data streams with concept drift.

Data mining approaches work well with large amounts of static data stored in system but does not address the problem of continuously flowing data. Data Stream mining is an archetype addresses the issues related to continuously arriving data. Handling and processing online data steams require single examination of data, fast processing with minimum space utilization, availability of results on the request of user (Lakshmi & Reddy, 2015). More sophisticated techniques need to be designed as the flowing data streams evolve over a period of time (Mehta, 2017).

Processing streaming data with concept drift is mostly applied in classification, clustering and for predictive analysis. This chapter introduces the techniques, algorithms and constraints while mining the evolving data streams which were designed and developed by well-known researchers. It is anticipated that the chapter will greatly help and provide a reference to researchers, practitioners and students interested in the emerging domain of evolving data streams.

## BACKGROUND

Online learning and designing new techniques for data stream mining is a challenge faced by researchers. In data streams data arrives rapidly and if not processed instantly it will be lost forever. Additionally, it is not practical to store this data in active storage media for longer period of time.

Extraction of knowledge from evolving data streams is becoming a key task for several researchers (Lakshmi & Reddy, 2010). Mining such evolving data streams have multitudinous applications with many daunting research issues. Handling concept drift in data streams can impact multidisciplinary domains (Khamassi et al, 2018). Before proceeding to further discussion, a study of the features related to data streams is presented.

### Features of Data Streams

The following features act as constraints over a model built for working with streaming data.

## Related Content

The Strengths, Weaknesses, Opportunities, and Threats Analysis of Big Data Analytics in Healthcare

Chaojie Wang (2019). *International Journal of Big Data and Analytics in Healthcare (pp. 1-14).*

www.irma-international.org/article/the-strengths-weaknesses-opportunities-and-threats-analysis-of-big-data-analytics-in-healthcare/232322

Development of a New Means to Improve the Performance of Self-Organizing Maps

Vijaya Prabhagar Murugesanand Punniyamoorthy M. (2022). *International Journal of Data Analytics (pp. 1-16).*

www.irma-international.org/article/development-of-a-new-means-to-improve-the-performance-of-self-organizing-maps/307065

Proficient Clustering algorithm for Wireless Sensor Networks

Nivetha Gopaland Venkatalakshmi Krishnan (2017). *Intelligent Multidimensional Data Clustering and Analysis (pp. 345-357).*

www.irma-international.org/chapter/proficient-clustering-algorithm-for-wireless-sensor-networks/172561

An Experimental Data of Lithium-Ion Battery Time Series Analysis: ARIMA and SECTRAL Analysis

Liming Xie (2021). *International Journal of Data Analytics (pp. 1-26).*

www.irma-international.org/article/an-experimental-data-of-lithium-ion-battery-time-series-analysis/285465

Different Approaches to Reducing Bias in Classification of Medical Data by Ensemble Learning Methods

Adem Doganer (2021). *International Journal of Big Data and Analytics in Healthcare (pp. 15-30).*

www.irma-international.org/article/different-approaches-to-reducing-bias-in-classification-of-medical-data-by-ensemble-learning-methods/277645