Stream Data Mining: A Survey

In this paper, author describes the overview of the growing field of Data Streams. To analyze the streams of data precisely, they cover theoretical basis as well. Also, they discuss the various techniques, which are used for mining data streams. The center point of this paper is to study the problems, which are involved in mining data streams. The methodology refers to summarization of the whole data set or selection of a subset of the incoming stream to be analyzed. The techniques used are Sampling, load shedding, sketching, synopsis data structures and aggregation. Sampling is the technique which can be made by a probability of the data item to be chosen. It is used for selection of the stream elements which are to be analyzed. Load shedding is the process which allows the data streams to be dropped overtime. The process has two steps. The methods involved in the sampling are Histograms, Wavelets and Micro cluster-based summarization. This paper includes different algorithms such as approximation algorithm, sliding window algorithm and output granularity algorithm as well, which are used for efficient utilization of time and space. Author examine and describes each of these algorithms in the way of analyzing data streams. These algorithmic ideas proved extremely useful and powerful for tackle a numerous problem in data streams. Several algorithms for extraction of knowledge from data streams were proposed in the domains of clustering, classification, Frequency counting, Time Series analysis and association. At the end, author concludes with a brief discussion of the big open problems and some promising research directions in the future in the area. Also, part of final discussion is that most of the current mining approaches utilize one passes mining calculations and few of them even locate the issue of drifting. The current methods produce approximate outcomes because of restricted memory. Examination in data streams is still in its beginning phase. Frameworks have been carried out utilizing these methods in genuine applications. On the off chance that the issues are tended to or settled and if more proficient and client friendly mining strategies are created for the end clients, certainly sooner rather than later data stream mining will assume a significant part in the business world as the data flows ceaselessly.