**GSOC 2013 proposal**

**Overview:** In an intensive care unit (ICU) there are multiple patients and each patient has multiple sensors emitting data continuously. For example, heart rate is generated once every two seconds whereas frequency of electrocardiography (ECG or EKG) is greater than 120Hz. ICUs generate large amount of streaming data that needs to be analyzed and visualized in real-time.

Though many analysis algorithms are used for retrospective data analysis, very few algorithms are implemented in a real-time framework. In this project, we will use a distributed streaming platform like Storm (<http://storm-project.net/>) or Esper (<http://www.espertech.com/products/esper.php>) to implement and incorporate multiple proven threshold, trend, and machine learning based algorithms to generate alerts and provide better knowledge of the patient physiologic state in real-time. This is the analytic part of the ICU decision support system when used in conjunction with visualization project described below.

Project Proposal

For this project I be looking at machine learning algorithms for streaming data. The large amounts of streaming data will be tackled by storm that will distribute the workload into clusters.

Storm is a distributed real time computation system that can be used with any programming language. Storm provides primitives (spouts and bolts) for transforming a stream that’s an unbounded sequence of tuples. For example, a stream of tweets into a stream of trending topics.

For this project I will need to gain understanding of data streams and drifting data. Data streams have data changing over time and will require data mining techniques that have adaptive mechanisms in order to maintain accuracy. The same data streams will need data mining models that can fit into limited memory, processing time and able to predict at any time.

I will divide my time to focus on :-

-Adaptive predictive modeling, which looks at classification and regression approaches for data streams. I will focus on the major type of approaches that are presented in academic papers for example having a single leaner with variable windows or fixed windows with instance weightings, ensemble with dynamic integration or adaptive fusion rules

-Clustering of streaming data. This looks at grouping at similar object together.I will look at clustering algorithms for example Stream K++ algorithms

-Machine learning problem is supervised learning

Learn a classifier form the data.

-Model :Bayesian Network Classifiers :learning BNCs from data: probablisitic graphical modeling

-learning scenario :on-line learning ,data arrives sequentially

flexabilty

scalable

real time and latency gurantees