

## Project Ideas for CSci 8980: Data Mining

---

**Spatial Data Mining.** Define the problem; what are its applications; how is it different from other data mining applications; what data mining techniques are used; what are the challenges of solving the problem; comparison between the different proposed solutions; suggestions for other techniques and improvements.

**Autocorrelation.** This is a characteristic of most spatial and temporal data which causes the values of physically adjacent data points to be closely related. For example, the temperature of a metal rod varies continuously from point to point within the bar and two points that are close will have almost the same temperature. Describe how this characteristic of spatial data affects clustering, classification and association rules.

**Multi-resolution analysis – spatial.** Analysis of the data at multiple spatial resolutions is very important since different patterns may reveal themselves at different scales. In image processing this is a common problem and various techniques, e.g., wavelets, are used to address it. Investigate this issue with respect to clustering, classification, or association rules.

---

**Sequential Patterns.** Define the problem; what are the challenges of solving the problem; what are the techniques used in solving the problem; comparison between the different proposed solutions; suggestions for other techniques and improvements.

**Detecting cyclical patterns.** Many time series and signal processing techniques are concerned with detecting cyclic patterns. However, data mining techniques for detecting sequential patterns, especially, sequential association rules don't have that capability.

**Multi-resolution analysis – temporal.** Analysis of the data at multiple temporal resolutions is very important since different patterns may reveal themselves at different time scales. To some extent the use of different windows sizes can help to deal with this problem.

---

**Markov Models.** Present a taxonomy for comparing the different applications of Hidden Markov Models. Describe what kinds of problems can be modeled using this method; how the technique works; how is it applied to a particular domain of application; what are the challenges; how those problems are solved; what are other applications.

---

**Data mining for Genomics, Bioinformatics.** Define the problem; how is it different from other data mining applications; what are the challenges of solving the problem; which data mining algorithms are used in solving the problem; comparison between the different proposed solutions; suggestions for other techniques and improvements. Focus on the spatial and sequential aspects of the problem.

**Data mining for Ecosystem Sciences.** Define the problem; how is it different from other data mining applications; what are the challenges of solving the problem; which data mining algorithms are used in solving the problem; comparison between the different proposed solutions; suggestions for other techniques and improvements. Focus on the spatial and sequential aspects of the problem.

---

**Application of Clustering on Web Usage Data.** Define the problem; why do we need to do clustering on Web usage data; what kinds of clusters are we looking for; what are its potential applications; what are the challenges faced; compare and contrast the different techniques used; suggestions for other techniques.

**Improving system performance with Web Usage Mining.** Web Usage Mining can be used to understand the behavior of Web traffic at a particular site. This in turn can be used to re-organize the structure of a Web site and to develop policies for Web caching, network transmission, load balancing, etc. Describe how this can be done; what are the important issues we need to look at; what are techniques people have used; how good are those techniques; what do you think can still be improved.

**Modeling of Web user profile.** Modeling of user behavior at a Web site enables dynamic presentation of Web content, tailored to the needs of users. Describe how Web Usage Mining can be used to achieve this goal; what are the challenges; how can it be solved; what kinds of data needs to be collected; how to build a good model; what data mining techniques can be (or has been) used.