**Aims and Background Material**

Big data are data sets which are so large or complex that traditional data processing or learning applications will lead to a poor performance. Some challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, etc. Big data causes computational difficulties and intrinsic statistical difficulties due to the data set being of large dimensions. This can cause overfitting, false structures, data isolation, etc. As data grows day by day, exploring different ways of dimension reduction is essential. The aim of my project will be to implement few generic/global fast projections that will reduce the dimensionality of a data set. Thus, the new data set can be used for traditional data processing or learning.

**Summary of Project Deliverables, Fall Backs & Extensions**

Project Deliverables:

* Explore different methods for dimensionality reduction and the mathematics behind each method
* Select 2 methods for further study
* Implement the methods
* Apply traditional data processing/learning (ex: K-nearest neighbours, SVM)
* Compare results of both methods
* Compare results before and after dimensionality projection

Fall Backs:

* Unable to handle the course modules and project simultaneously
* Insufficient information required to progress from a given stage
* Insufficient knowledge to understand areas in research papers

**Summary of Risks**

* Lack of understanding the mathematics behind the implementation