<table>
<thead>
<tr>
<th>Course Code</th>
<th>MH4510</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Title</td>
<td>Statistical Learning and Data Mining</td>
</tr>
</tbody>
</table>
| Pre-requisites | MH2500 - Probability & Introduction to Statistics  
MH3500 - Statistics  
MH3510 - Regression Analysis  
MH3511 - Data Analysis with Computer |
| Contact Hours | 39 hours of lecture+13 hours of tutorial/lab |

**Course Aim:**
This course gives an overall view of the modern statistical/machine learning techniques for mining massive datasets.

**Intended Learning Outcomes:**
1. resolve data mining problems with various modern statistical techniques;
2. summarise the strengths and shortcomings of different techniques;
3. evaluate learning methods statistically and recommend the optimal one for applications;
4. implement the modern statistical techniques with statistical software such as R.

**Course Content:**
1. Optimal decision rules and K-nearest neighbors methods;
2. Linear models for regression;
3. Generalized linear models for classification;
4. Cross-validation and bootstrap methods;
5. Ridge Regression and LASSO;
6. Artificial neural networks;
7. Classification and regression trees and ensemble methods;
8. Clustering methods;
9. Advanced topics.

**Reading and References:**

**Textbook:**

**Reference:**


*School of Physical and Mathematical Sciences*  
*Division of Mathematical Sciences*