F20ML Statistical Machine Learning

COURSE DETAILS

Course Code: F20ML
Full Course Title: Statistical Machine Learning
SCQF Level: 10
SCAF Credits: 15
Available as Elective: No

DELIVERY LEVEL

Undergraduate: Yes  Postgraduate Taught: No  Postgraduate Research: No

COURSE AIMS

To provide students with an in-depth introduction to data mining and machine learning and to the mathematics underpinning these techniques. Topics will include generative and discriminative approaches, classification, clustering, regression, and supervised and unsupervised learning. Students will develop practical experience by using the WEKA software to apply a range of machine learning algorithms to benchmark datasets.

LEARNING OUTCOMES – SUBJECT MASTERY

- An understanding of the fundamental concepts in data mining and machine learning.
- An understanding of the mathematics underpinning the data mining and machine learning techniques.
- Critical awareness of the appropriateness and performance of the different techniques, as well as the relationships between them.

LEARNING OUTCOMES – PERSONAL ABILITIES

- Rational problem identification and definition.
- Critical analysis and solution selection
- Demonstrate the ability to learn independently.
- Manage time, work to deadlines, and prioritise workloads.
- Use appropriate computer software to process data.
- Present results in a way that demonstrates a good understanding of the technical and broader issues of data
mining and machine learning.

SYLLABUS

**Basic Concepts**: classification, clustering, regression, supervised and unsupervised learning.

**Generative Models**: probabilistic graphical models; cluster analysis (k-means, expectation-maximisation, mixture models, hierarchical models); regression analysis.

**Discriminative Learning**: Instance-based learning and decision tree learning; artificial neural networks (perceptron, MLPs, back propagation, introduction to deep learning architectures); maximum entropy models; support vector machines; ensemble learning (bagging, boosting, stacking, random forests).

LOCATION AND ASSESSMENT METHODS

<table>
<thead>
<tr>
<th>Edi</th>
<th>SBC</th>
<th>Ork</th>
<th>Dub</th>
<th>Malay</th>
<th>IDL</th>
<th>COLL</th>
<th>ALP</th>
<th>OTH</th>
<th>Method</th>
<th>Weight</th>
<th>Exam Mins</th>
<th>Type</th>
<th>Diet</th>
<th>Synoptic Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Examination</td>
<td>60</td>
<td>120</td>
<td>Assessment</td>
<td>Semester 1</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Coursework</td>
<td>40</td>
<td></td>
<td>Assessment</td>
<td>Semester 1</td>
<td></td>
</tr>
</tbody>
</table>