Module title: Machine Learning and Big Data in Finance

Please note this module description is preliminary and is subject to change.

Module code: Providing School/Department: ICMA Centre
Level: 7 Number of credits: 20
Term(s) in which taught: Spring Number of ECTS credits: 10
Module convenor: TBC

Summary module description
In this module you will learn how machine learning techniques borrowed from artificial intelligence can be used to solve common big data problems in finance. We will first explore the issues related to the collection, organisation and visualisation of large sets of structured and unstructured data. With the use of Python we then explore ways in which a computer can be trained to recognise patterns in the data and its popular finance applications. For instance, we will look at stock price forecasting, company default prediction and market sentiment analysis.

Aims
The module focuses on (1) issues facing big data handling (2) high level description of distributed storage and processing of big data (Hadoop) (3) retrieval, organisation and cleaning of structured and unstructured data (4) visual analysis of a dataset (5) common machine learning techniques such as logistic regression, decision trees, K-nearest neighbours, k-means clustering, principal component analysis and deep learning tools like neural networks (6) finance applications.

Intended learning outcomes:

Assessable learning outcomes:
By the end of the module it is expected that students will:
Understand how big data and artificial intelligence are changing our lives and creating business opportunities
Be familiar with the main issues in distributed storage and processing of big data
Understand the basic techniques for the collection and cleaning of large structured and unstructured data;
Understand the need for a rigorous data science approach and the concepts of training data, validation data and testing data;
Be able to build machine learning models and interpret the models in terms of their structure and accuracy;
Understand how big data and machine learning can be used to solve old and new problems in finance

Additional outcomes:
The module will use the industry standard Python programming language and will build on the programming skills developed in Part 1.

Outline content
1. Big data – a global multi-sector view
2. Distributed storage and processing of big data (Hadoop)
3. Structured and unstructured data collection, organisation, storage and cleaning
4. Artificial intelligence, machine learning, deep learning
5. Linear and logistic regression models in Python and finance applications
6. Decision Tree Models in Python and finance applications
7. K-nearest neighbours and k-means clustering in Python and finance applications
8. Principal component analysis in Python and finance applications
10. Big data and machine learning: case studies

Global context (where appropriate):
The module covers industry standard techniques using international datasets. The concepts are applied in investment banks, central banks, hedge funds and asset management firms worldwide.

Brief description of teaching and learning methods:
The core theory and concepts will be presented during lectures. Problem sets will be solved in workshops.

Contact hours:

<table>
<thead>
<tr>
<th></th>
<th>Autumn</th>
<th>Spring</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures</td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>Seminars</td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Tutorials</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project supervision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practical classes and workshops</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervised time in studio/workshop</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fieldwork</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>External visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work-based learning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guided independent study</td>
<td></td>
<td></td>
<td>170</td>
</tr>
<tr>
<td>Placement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year abroad</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total hours by term</td>
<td></td>
<td></td>
<td>200</td>
</tr>
<tr>
<td>Module total hours (10 credit module = 100 hours)</td>
<td></td>
<td></td>
<td>200</td>
</tr>
</tbody>
</table>

Summative Assessment Methods (%) - work which contributes towards the overall module mark:

<table>
<thead>
<tr>
<th>Method</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written assignment, including essay</td>
<td></td>
</tr>
</tbody>
</table>
Summative assessment - Examinations:

Summative assessment - Coursework and in-class tests:
Students will be asked to complete a report (40%) in week 2 of the summer term and two in class multiple choice tests (30% each) in weeks 7 and 11 of the spring term.

Formative assessment methods:
Seminar questions are assigned for each class. The seminar leader will facilitate discussion and offer feedback.

Penalties for late submission:
Penalties for late submission on this module are in accordance with the University policy. Please refer to page 5 of the Postgraduate Guide to Assessment for further information:
http://www.reading.ac.uk/internal/exams/student/exa-guidePG.aspx

Assessment requirements for a pass:
50% weighted average mark

Reassessment arrangements:
Reassessment of individual report

THE INFORMATION CONTAINED IN THIS MODULE DESCRIPTION DOES NOT FORM ANY PART OF A STUDENT'S CONTRACT.