ACM Data Science Task Force Course Example

Machine Learning for Data Analytics *(CS985)*

University of Strathclyde, UK
Dr Marc Roper, Dr Leif Azzopardi
marc.roper@strath.ac.uk

Knowledge Areas that contain competencies (knowledge, skills, and dispositions) covered in the course

Knowledge Area	Total Number of Contact Hours
Machine Learning (ML)	18
Programming, data structures and algorithms (PDA)	2
Big data systems (BDS)	17
Professionalism (PR)	3

Where does the course fit in your undergraduate Data Science curriculum?

This course is given in the second semester of a Masters course in Data Science, for students who have no prior qualifications in Data Science.

Is this course from or used in other curricula/majors?

It is available for students on other Masters programs

What is covered in the course?

The aim of this course is provide an overview of the concepts, methods and tools needed to implement machine learning applications - essentially programs that are capable of learning from data. The class assumes no knowledge of machine learning but some familiarity with the Python programming language (or a willingness to learn it very quickly).

The class consists of two parts:

- Part 1 (weeks 1-5) covers the fundamentals principles of machine learning, introduces the machine learning process, and covers a range of classification and regression algorithms including linear and logistic regression, decision trees, support vector machines, random forests and also unsupervised learning (clustering).
- Part 2 (weeks 6-11) covers more advanced topics, in particular deep learning, neural networks and TensorFlow

What is the format of the course?

Normally this one semester course would be face-to-face: there would be 20 hrs of lectures accompanied by 20 hours laboratory work. An additional 160 hrs of private study is envisaged. Given the problems caused by Covid, it is currently to be taught remotely.

How are students assessed?

There is a two-hour examination worth 50% plus two assignments:

Assignment 1 (25%) This is done in groups of 4. The aim is to put into practice the concepts covered, to apply them to a real dataset, and to demonstrate an ability to use Python to carry out machine learning tasks. Two problems are to be addressed: a regression problem which aims to predict the popularity score of a song; and, a classification problem which aims to predict the top genre to which a song belongs

Assignment 2 (25%) Students work in groups of maximum size four. The aim of the assignment is to demonstrate the ability to use Python to carry out machine learning tasks using deep learning with Keras. The data to be working on comes from two sources (Twitter and Kannada-MNIST). The two problems that you will need to be solved are:

- A classification problem where you need to predict the sentiment of tweets.
- A classification problem where you need to predict the hand-written digits.

Course tools and materials

The course closely follows "Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems" by Aurelien Geron.

Why do you teach the course this way?

Students on this course will have obtained a degree previously, but not in Data Science. Consequently the teaching reflects a Masters approach.

Body of Knowledge coverage

KA	Sub-domain	Competencies Covered	Hours
ML	General, Supervised learning, Unsupervised learning, Deep learning	• Understand the key concepts underpinning the various models for machine learning • Apply machine learning algorithms. • Appreciate the breadth and utility of machine learning methods, compare and contrast them, select appropriate methods for specific problems.	18
PDA	Programming, Data structures, Algorithms,	• Write clear and correct code in a programming language in Python.	2

	Algorithmic thinking and problem solving	● Use standard libraries for data science	
BDS	Problems of scale, Techniques for big data applications. Software support for big data systems	● Instil confidence in dealing with the problems of Big Data ● Provide familiarity with a range of skills that may be used in the implementation of Big Data applications	17
PR	Teamwork, Communication, privacy and confidentiality, Ethical considerations	The two major assessments are carried out in team settings. Accordingly professionalism competencies are addressed in this way.	20