

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
BD001	Introduction to Data Science	Dr Jochen Leidner, Thomson Reuters	<b>27/07/2017</b> 09.00-17.00 <b>28/07/2017</b> 09.00-17.00	Introductory 2 days (12 hours)	<p>The goal of this compact course is to give participants a first gentle introduction and solid conceptual grounding in what has been called ‘data science’, i.e. experimental work that is data-driven and empirical. The focus is on methodology, defining an experimental protocol, devising hypotheses, thinking about measuring success, but also on more practical approaches like basic machine learning methods (both supervised and unsupervised) and natural language processing approaches (like part-of-speech tagging, named entity recognition/classification/resolution, and parsing) and the introduction to popular tools. The course also demonstrates some practical applications of the techniques shown, and deepens the students’ skills via practical exercises.</p> <p><b>Contents</b> Structured and unstructured data; profiling data sets; sampling, pre-processing &amp; cleaning; hypothesis testing; descriptive and predictive analytics; clustering and classification; k-means and agglomerative clustering; naive Bayes; linear and logistic regression; HMMs &amp; Viterbi search; neural networks and deep learning; Web crawling &amp; mining; practice tools, techniques &amp; methodologies; KDD, SEMMA, CRISP-DM &amp; D2V; experimental protocol; evaluation measures; inter-rater agreement; applications; data science economics; value creation; visualization &amp; presentation; planning your data science project; data science &amp; ethics.</p>	This 2-day course targets advanced undergraduate students and master’s students in computer science	<ul style="list-style-type: none"> <li>· Basic programming knowledge in any language/algorithms &amp; data structures</li> <li>· Python skills</li> </ul> <p>Own laptop required</p>

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
BD002	Hadoop Fundamentals for the Cloud	Dr Martin Fleury, University of Essex	<b>28/07/2017</b> 09.00-17.00	Intermediate, 1 day (6 hours)	<p>Apache's open-source Hadoop version 2 provides MapReduce processing as a cloud service. The cloud user can submit pre-written programs in the MapReduce format to process giga-, tera-, peta-bytes of data. For example, Facebook has a Hadoop Distributed Filing System (HDFS) cluster of 100 petabytes. As such Hadoop is a key software technology for running Big Data applications.</p> <p>In this course, we introduce the cloud concept followed by the contributors to Hadoop 2, such as Zookeeper, Chubby, Google Filing System, Google's Big Table, and Kerberos. Alongside these contributors, the key concepts will be introduced such as eventual consistency, data replication, and consensus algorithms.</p> <p>A hands-on computer laboratory will complement the material, in which participants will configure and compile Hadoop programs to run on test applications. The Hadoop programs will run on the VirtualBox cloud-environment emulator from Oracle. Using VirtualBox, software developers trial and debug Hadoop applications before transferring to an actual cloud.</p> <p>Researchers will also find this course of interest, as, for example, the Blast algorithm for searching nucleotide databases is suitable for Hadoop implementation.</p>	Background in computer science and/or social science and/or business	Basic knowledge of computer operating systems, especially Linux/Unix.
BD003	Introduction to Natural Language Processing	Dr Diana Maynard, University of Sheffield	<b>31/07/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	Natural Language Processing (NLP) is a key component for all kinds of text analytics, including information extraction, social media analysis, sentiment analysis, semantic web technologies, information summarisation, question answering and much more. NLP provides the means to extract meaning from text	Participants with any background are welcome: it is suitable for participants from all disciplines who have an interest	No pre-existing knowledge of NLP or programming is required for this module. <b>Participants need to bring their</b>

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					rather than just statistical information. In this tutorial we will introduce the basic concepts of NLP, starting with simple text pre-processing techniques such as tokenisation and part-of-speech tagging, and moving on to more complex tasks such as term extraction, entity recognition and information extraction. The techniques will be demonstrated using GATE, one of the most widely used toolkits for performing all kinds of NLP tasks, and which is freely available and open source. GATE includes not only its own text processing components, but also includes a number of popular third party NLP components, all of which participants will be able to experiment with during the tutorial with hands-on exercises.	in text analysis. This course is a compulsory pre-requisite for those attending the Sentiment Analysis course.	<b>own laptops</b> for the practical sessions, and to download and install the GATE software in advance
BD004	Practical text analytics and sentiment analysis from social media	Dr Diana Maynard, University of Sheffield	<b>01/08/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	This tutorial will introduce the concepts of social media and sentiment analysis from unstructured text. It will first introduce the concept of social media analysis, showing how this form of noisy text requires different solutions from traditional text analysis, with practical examples and exercises showing how this can be achieved. This leads into the more specialised task of sentiment analysis: the problem of extracting opinions automatically from text. It will cover both rule-based and machine learning techniques, provide some information on the key underlying NLP and text analysis processes required, and look in detail at some of the major problems and solutions, such as detection of sarcasm, use of informal language, spam opinion detection, trustworthiness of opinion holders, and so on. The techniques will be demonstrated with real applications developed in GATE, an open-source language processing toolkit. Hands-on exercises and relevant materials will be provided for	Participants with any background are welcome: it is suitable for participants from all disciplines who have an interest in social media and sentiment analysis. However, <b>participants must have attended the Introduction to Natural Language Processing course unless they are already familiar with the GATE toolkit.</b>	No pre-existing knowledge of NLP or programming is required for this module. However, the participants must have attended the Introduction to Natural Language Processing course unless they are already familiar with the GATE toolkit. <b>Participants must bring their own laptops</b> for the practical sessions, and to download

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					participants to try out the applications, and to experiment with building their own simple tools.		and install the GATE software in advance.  Own laptop required
BD005	Practical methods for discrete optimisation	Dr Daniel Karapetyan, University of Essex	<b>04/08/2017</b> 09.00-17.00	Advanced, 1 day (6 hours)	<p>Discrete optimisation is the core of many decision support systems; whenever we need to produce a schedule or timetable, or find the best route, or assign resources, we face an optimisation problem which is usually hard to solve. In this course we will discuss modern approaches that do not require significant investment of time in algorithm development but still allow to tackle real-world problems. We will cover the following topics: what is optimisation and how is it related to decision support/making, off-the-shelf solvers (mixed integer programming), branch and bound algorithms, simple heuristics, metaheuristics (stochastic local search and population based methods), and practical aspects of algorithm engineering. The lectures will be intermixed with lab exercises to get hands on experience. By the end of the day you will know the main concepts in modern discrete optimisation. You will also have an idea how to choose the right approach for a given problem.</p> <p>During the exercises, you will implement several optimisation algorithms in Java, though building blocks will be provided, so you will need to write only relatively simple code.</p>		Computer science or mathematics background is expected (basics of graph theory, background knowledge of simple data structures and standard algorithms). To participate in lab exercises, one will need Java programming skills sufficient to write simple code, though solutions will be provided
BD006	IoT Application on Environment and Health Monitoring	Dr Sefki Kolozali/Nazli Farajidavar, King's College	28/07/2017 09.00-17.00	Introductory, Intermediate 1 day (6 hours)	Overview of IoT, the Semantic Web, and data analysis techniques. The course will highlight the Semantic Sensor Web technologies to deal with heterogeneity issues, and detail different time	Computer science and Machine learning	N/A

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
		London, University of Oxford			series analysis techniques for sensory devices. The course will also present machine learning applications on IoT data for extracting high-level information along with a workshop session and examples of practical implementations.		
BD007	Techniques for securing IoT devices in a cloud environment - CANCELLED	Dr Gareth Howells, University of Kent	27/07/2017 09.00-17.00	Introductory, 1 day (6 hours)	<p>The course will introduce a suite of technologies supporting the identification of IoT devices and cloud based services based on the derivation of digital signatures based on their behavioural characteristics, a technology termed ICMetrics. Specifically, ICMetrics possess the following significant potential:-</p> <ul style="list-style-type: none"> <li>Secure communication from IoT devices and cloud based services via the direct generation of digital signatures and encryption keys from the internal behavioural characteristics of software and hardware associated with the device.</li> <li>Prevention of unauthorised access to cloud services or IoT devices.</li> <li>Prevention of the fraudulent cloning or imitation of such devices.</li> <li>Implicit detection of tampering of the software or hardware associated with the device or service.</li> </ul> <p>The course will introduce basic Pattern recognition techniques such as Bayesian classifiers incorporating multi-variate Gaussian distributions and cryptographic techniques such as Diffie-Hellman key exchange and RSA encryption which are combined together to form the secure Icmetric key exchange environment.</p>		
BD008	Network analysis: from models to	Dr Giovanni Stracquadano,	26/07/2017 09.00-17.00	Introductory/ Intermediate,	From social networks to cell metabolism, many phenomena are the result of interactions	Science, mathematics,	Basic knowledge of statistics and

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
	community detection	University of Essex		1 day (6 hours)	<p>between many players. Network science provides a robust framework to model, study and predict the behaviour of these complex systems using algorithms on graphs.</p> <p>This course will focus on identifying structural patterns in networks using different mathematical models, i.e. random graph, small word and preferential attachment models. Then, we will discuss algorithms to identify communities in networks, starting from modularity maximization to stochastic block models.</p>	biology, social science	algorithms on graphs
BD009	Data science for public policy	Professor Slava Mikhaylov, University of Essex	<b>31/07/2017</b> 09.00-17.00	Introductory/ Intermediate, 1 day (6 hours)	<p>Public sector organisations increasingly express interest in using data science capabilities to develop and deliver policy, and generate efficiencies in high uncertainty environments. While public organisations successfully utilize standard data sources (e.g. surveys), there are many opportunities in generating additional value for organisations from large volumes of unstructured data (e.g. textual data from consultations and consumer feedback) and new forms of data (e.g. from smart energy meters). This is an applied course, where we build on the material covered in other modules and focus on the application of these concepts in the specific context of public policy analysis.</p> <p>The course is interactive, with a hands-on computer lab, in which participants will work through three analytical projects covering the application of machine learning techniques (topic models, clustering, predictive modelling) to the policies at the EU and UK levels.</p> <p>We will also focus on the critical issues of reproducibility of analytical results using R Notebooks and presentation of results to stakeholders with R Markdown.</p>	Public administration, public management, business, computer science/maths/ social science	<p>Basic knowledge of R</p> <p>A good introduction to statistical software is “R for Data Science” by Golemund &amp; Wickham, available online from: <a href="http://r4ds.had.co.nz">http://r4ds.had.co.nz</a></p>

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
BD010	Visual object recognition and tracking	Professor Ales Leonardis, University of Birmingham	<b>26/07/2017</b> 09.00-17.00	Intermediate 1 day (6 hours)	Computer vision provides essential tools that enable us to automatically deal with large image and video datasets. This course will primarily focus on visual object recognition and tracking, which are at the core of computer vision research. After introducing the basic concepts, the course will provide an overview of some of the most commonly used algorithms, both for object recognition and object tracking, and discuss various approaches for their performance evaluation, including visual challenges		MATLAB, Basic understanding of computer vision concepts and image processing algorithms
BD011		Dr Audrey Guinchard, Professor Anthony Vickers University of Essex	<b>31/07/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	This session aims to introduce the current EU and UK data protection regime and the changes to be brought in by the future General Data Protection Regulation applicable in May 2018, in spite of Brexit. Furthermore, the session will present and allow for discussion of the specific challenges big data analytics bring, especially in light of the reports published by various data protection regulators on big data both at UK and EU levels. Special attention will be given to security requirements in data protection law. The last two hours (with both speakers) will introduce the ethical issues arising from Big Data and present the correlative legal issues that may arise in light of Data Protection legislations and of criminal law. Torts and contracts will not be covered.		
BD012	Encounters with big data: An introduction to using big data in the social sciences - FULL	UK Data Service	<b>31/07/2017</b> 09.00-17.00 <b>01/08/2017</b> 09.00-17.00 <b>02/08/2017</b> 09.00-17.00 <b>03/08/2017</b> 09.00-17.00	Introductory, 5 days (30 hours)	This 5-day course run by the UK Data Service will introduce key concepts and discussions around using big data in the social sciences, and introduce attendees to Apache Hadoop, an open source framework for analysing big data. This course will focus on quantitative data and will not cover in any detail text, social media, audio or other forms of non-numeric data	Aimed at experienced researchers, statisticians, or data managers	Experience using quantitative research data in the social sciences,  A good understanding of

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
			04/08/2017 09.00-17.00		<p><b>Day 1: Introducing big data for the social sciences</b></p> <p>On day 1 we introduce concepts around big data and the challenges of working with it. We reiterate that we should still assemble data with a purpose and hypothesis in mind. The difference with big data is that we are not always as 'in control' of the data source. We give some examples of what social scientists are doing with big data and ask participants to undertake some group work to design an experiment to determine a national statistic based on web-sourced open data.</p> <p>Then we move onto thinking about structures of data for social scientists, and provide an example of how longitudinal datasets can be structured. Following this we will look at ethics and rights in big data, providing example of the 5 Safes approach for robust data governance, and explore how data available from the internet may pose additional hurdles for ethics and disclosure risk. In the final session we introduce Hadoop as a 'big data' solution and set out some of its key features.</p> <p><b>Day 2: Manipulating data using Hive</b></p> <p>On day 2 we will introduce you to the Hadoop sandbox for storing and manipulating big data and several of the included components. Through a combination of presentations, demos and hand on exercises we will cover:</p> <ul style="list-style-type: none"> <li>● Loading big datasets into the Hadoop file system and processing them using Hive</li> <li>● Using Hive Query Language (HiveQL) to examine the contents of the datasets and to 'slice' and 'dice' the dataset into a smaller</li> </ul>		<p>statistical methodology and concepts like standard error and standard deviation,</p> <p>Competence in writing commands in a statistical computing environment like Stata, R or SPSS.</p>



Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					<p>datasets which can be used by desktop applications</p> <ul style="list-style-type: none"> <li>• Creating internal and external tables and more advanced structures of partitioned tables</li> <li>• Accessing Hive using a variety of methods including Zeppelin and the Ambari views.</li> </ul> <p><b>Day 3: Manipulating and analysing data using R and SparkR</b></p> <p>On day 3 we introduce Apache Spark, a high performance, distributed computation engine designed for handling big data. In particular, we demonstrate how we can harness the power of Spark from R using the SparkR package. We conclude by giving an overview of R's capabilities for producing spatial visualizations such as choropleth maps.</p> <p>Day 3 will cover:</p> <ul style="list-style-type: none"> <li>• Introduction to R and Spark</li> <li>• Exploratory data analysis using SparkR</li> <li>• Linear models using R and SparkR</li> <li>• Creating maps in R</li> </ul> <p><b>Day 4: Working with your data</b></p> <p><b>Day 4: Tools and techniques for dealing with external data; Projects</b></p> <p>The first sessions will cover tools and techniques for getting with and converting external data from APIs, and an overview of end-to-end process tools on Hadoop.</p> <p>External data obtained from the Internet either comes as complete files often in structured format such as csv or via web services or APIs in which case you need to be able to construct a valid API call and know how to interpret the results which will be in JSON format. We cover creating simple scripts to download a single or a set of files as well as looking at APIs and</p>		

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					<p>demonstrating downloading data from an API and extracting useful data from the JSON stream. Participants will then work in groups on projects they design to practice using some of the tools and techniques covered.</p> <p><b>Day 5 Project presentations, data papers and using GitHub</b></p> <p>In the morning each project group will finalise their project and briefly present their work. In the afternoon, we will cover publishing and sharing data : replication, peer review, data papers and code sharing.</p> <p>The first presentation will introduce the requirement for robust replication of findings in published. Articles, now imposed by an increasing number of journals in the sciences and social sciences. These expectations for transparency can be met by scholars by appreciation and knowing how best to publish high quality raw (or processed data), methodology and code. We will discuss what this means for publishing in the social sciences and how big data fit in. We look at the role of the ‘data paper’ to help elucidate how data were constructed, compiled and processed, and also showcase the value of data beyond the original research. We show what a data paper could look like and ask participants to consider how they might turn something in their own research into a data paper. As a group we consider the problems of describing data with uncertain provenance as may be the case with big data - how can we ensure it can be replicated, or can we?</p> <p>Finally, we present an example of how to set up your own Github repository, for example to share code and syntax, and how to build on someone else’s published code</p>		

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
BD013	Human rights in the era of big data and analytics	Dr Daragh Murray and Ms Cath Kent	<b>24/07/2017</b> 11.00-17.00	Introductory, 1 day (4.5 hours)	This course will examine the impact of big data and analytics on human rights protections. It is intended to provide an insight into how big data and analytics affect the enjoyment of human rights. We will look at the most relevant provisions of international human rights law, discuss how these are affected by big data and analytics – either positively or negatively – and begin a discussion on how human rights can be taken into consideration in your work.	This course is open to all interested individuals	None
BD014	From data science to decision making	Dr Spyros Samothrakis, University of Essex	<b>02/08/2017</b> 09.00-17.00 <b>03/08/2017</b> 09.00-17.00 <b>04/08/2017</b> 09.00-17.00	3 days (18 hours)	The aim of this module is pull together all the different strands of modern data science and artificial intelligence. We will explore diverse applications that range from website optimisation to generating images to game playing, all within a unifying framework. We will bring together topics from Neural Networks, Bandits, Causality and Reinforcement Learning, with the explicit aim of creating a unified understanding of the topics that transcends current trends. The module has a strong practical component and requires good python programming skills		Good programming skills in Python, some familiarity with mathematics
BD015	Introduction to R	Professor Leo Schalkwyk, Dr Nicolae Zabet, University of Essex	<b>24/07/2017</b> 11.00-17.00 <b>25/07/2017</b> 09.00-17.00 <b>26/07/2017</b> 09.00-17.00 <b>27/07/2017</b> 09.00-17.00	Introductory, 4 days (22.5 hours)	R is an interactive computing environment and programming language designed for statistical analysis and graphics. Extensions to the basic capabilities of R are straightforward to produce and share with others. It is widely and increasingly used in many Big Data fields of research including bioinformatics. Because of its power and flexibility, R is more demanding to learn than traditional statistical packages but rewards some initial effort. This course is based tested material that we have been using for nearly 10 years to help research students, postdocs and faculty get started in their own data analysis, and is refined each time based		

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					<p>on feedback. It is aimed at people who may have little or no programming experience.</p> <p>The course will emphasize the fundamentals of the R language in an intensive format where each student has a computer and 50% of the time is spent on practical exercises, and will include a special module on big data techniques.</p> <p><b>Note</b> – participants on this course must attend all 4 days</p>		
BD016	Agent-based modelling of social systems	Dr Peter Barbrook-Johnson, Dr Kavim Narsimhan University of Westminster	<b>02/08/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	<p>This course will introduce participants to agent-based modelling (ABM). ABM is a computational modelling approach in which we build simulations of social and policy processes in an attempt to understand better those processes. We can also use these simulations to ask ‘what-if?’ type questions (i.e., to explore policy scenarios). In this course we will explore what ABM is using recent social and policy examples, and get hands on, building a simple model in NetLogo, and playing with an actual policy model. No prior knowledge is needed. Participants will leave the course with a clear understanding of what ABM is, and when (and how) it can be a powerful approach to use</p>	PhD students and ECRs in social science disciplines. No specialist knowledge/ expertise needed	None
BD017	Machine learning with Mahout	Richard Skeggs, ESRC Business and Local Data Research Centre, University of Essex	<b>04/08/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	<p>This is an introduction into the use of machine learning algorithms supported by the Apache Mahout framework. The class will concentrate on what problems can be solved using Mahout before looking at the common classifiers used by Mahout to achieve those objectives. Finally the class will look at building some simple working examples to see Mahout in practice.</p>		Knowledge of the Java programming language is essential. Some statistical knowledge will be useful but not essential.
BD018	Hands-on Apache Spark: - “multifaceted”	Dr Ivan Palomares Carrascosa,	<b>01/08/2017</b> 09.00-17.00 <b>02/08/2017</b>	Intermediate, 2 days (12 hours)	<p>The Apache Spark open source framework has gained significant popularity in the Big Data universe: it enables writing large-scale data</p>	Computer Science, or Statistics/	Basic knowledge of a computer programming

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
	data processing and machine learning	Queens University Belfast	09.00-17.00		analytics applications - from distinct data sources and repositories – that perform much faster than classical cluster architectures. Likewise, its compliance with several popular programming languages makes it an accessible tool to a wide community of developers and data scientists. This hands-on course firstly introduces Apache Spark engine for data analytics and unveils its core characteristics and functions. Secondly, the course explores two popular Spark libraries for data processing, showing their versatility for writing analytics applications that process data stemming from multiple sources and/or with different nature (structured data and streaming data processing). Finally, the course introduces some well-known Machine Learning (ML) techniques such as classification, regression and clustering techniques; and describes how to utilize Spark to build scalable ML applications based on them.	Mathematics/ Business audience with a basic understanding of computer programming and Databases	language (ideally one among Python, Scala or Java), basic database concepts and introductory statistics.
BD019	Big Data and Financial Analytics	Dr Nikos Vlastakis, University of Essex	<b>03/08/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	Given the current information rich business environment, as well as the ever increasing available computing power and potential for data collection, it is not surprising that many firms are interested in big data. This is particularly true for firms in the financial sector, where the overabundance of data raises new challenges that include capture, curation, storage, search, sharing, transfer, and of course data analytics and visualization. The purpose of this course is to provide the participant with an understanding of data analytic approaches in finance. The first part covers high frequency trading and predictive analytics. The second part will concentrate on the application of data analytics in risk modelling, corporate finance, fraud and personal finance	Some background in statistics/ mathematics/ econometrics is desirable but not essential	None
BD020	Introduction and	Ragini Gokhale,	<b>24/07/2017</b>	Intermediate,	<b><i>Introduction to Cassandra</i></b>	Administrators,	Understanding of

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
	Data Modelling with Cassandra	University of Essex	11.00-17.00 <b>25/07/2017</b> 09.00-17.00	2 days (10.5 hours)	<p>Understand challenges faced to scale relational databases and how Cassandra, the no SQL database, provides the mean to resolve them</p> <p>Understand fundamentals of Cassandra, its architecture and how data is stored.</p> <p>Install and configure Cassandra Internal architecture by understanding the read, write path and compaction CQL Partition and Clustering Replication and Consistency</p> <p><b>Data modelling with Cassandra</b></p> <p>Data modelling techniques for Cassandra deployment Principles, methodologies, design techniques and optimisations Test/Understand data modelling use cases Relationship Keys Hierarchy and conceptual model Application workflow and access patterns Mapping rules and patterns Usecases</p>	Architects, Developers	<p>Relational Database (SQL, mySQL)</p> <p>Basic programming knowledge</p> <p>Exposure to Linux, Java, Apache</p> <p>Concepts of distributed systems</p>
BD021	Doing data science	Harry Powell, Barclays PLC	<b>01/08/2017</b> 09.00-17.00	Intermediate, 1 day (6 hours)	<p>So how do you <i>do</i> data science? Not just how do you write a line of code, or how do you tune a neural network, but how do you think about difficult real world problems in an abstract way and solve them using pattern recognition, distributed computation and automation?</p> <p>In this course we will take four real data science applications that we have built at Barclays. We will explore the modelling, methodology and design decisions in addressing these problems and the challenges posed by implementing machine learning applications in a big data environment.</p> <p>The focus will be on imaginative problem solving in small teams and in engaging critically but positively with other people's ideas. Hopefully we'll come up with something new and ground-breaking!</p> <p>There will be no computers in this class, just</p>	The problems will require a reasonable understanding of computer science, maths, statistics and machine learning	None

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					loads of blank sheets of paper and a lot of scribbling. You will be expected to present your ideas to the class.		
BD022	Introduction to data visualisation using R	Dr Aris Perperoglou, University of Essex	26/07/2017 09.00-17.00	Introductory/ Intermediate 1 day (6 hours)	<p>Graphics reveal data. In this course we will illustrate how to perform graphical data analysis to enhance your own understanding of hidden patterns and relationships, and how to present data in a way that yields insight and communicates important information. We will be using Rstudio as the toolset and advanced modern packages such as ggplot, shiny and googleVis.</p> <p>The course will cover the topics:</p> <ul style="list-style-type: none"> <li>- Data visualisations: common mistakes</li> <li>- Using R for basic plots</li> <li>- Advanced graphics using ggplot</li> <li>- Plotting categorical and continuous data</li> <li>- Identifying relationships</li> <li>- Multivariate data</li> <li>- Interactive documents and plots with shiny</li> <li>- Google charts in R</li> </ul>	Researchers who deal with quantitative data	Basic understanding of statistics, R skills
BD023	Designing research for your data	Tarek Al Baghal, University of Essex	<b>24/07/2017</b> 11.00-17.00	Introductory, 1 day (4 hours)	This course will provide an introduction on how to design quality research projects for use with generated data in order to produce meaningful outcomes. With the continually increasing amount of data available, the variety of research possible also grows exponentially. However, the existence of data does not ensure the ability to produce meaningful analysis and results. We will cover the development of research questions, goals and desired outcomes, and how the data available can address these questions and address these goals. Important aspects of the data we discuss in relation to research design are the structure of data, identifying possible useful indicators in the data, and what additional	Open to any research discipline, but instructors and examples present from a social science perspective	None

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					measures can be created from the data. Equally important, we will discuss how the data can limit research projects, such as not having indicators related to the research objectives or having errors in the data, and how to address these issues. Examples of this research process will be given using actual ongoing research projects, including using varied administrative data generated from a non-profit organisation and the use of Twitter data generally and in combination with survey data		
BD024	Principles of big data storage and processing systems	Dr Nikos Ntarmos, University of Glasgow	<b>27/07/2017</b> 09.00-17.00	Introductory/ Intermediate 1 day (6 hours)	Enterprises of all sizes are struggling to cope with what can only be described as a deluge of data; as a matter of fact, business data is estimated to double every 1.2 years or less, meaning that almost every year we produce more data than in all previous years combined. In the face of what is now called Big Data, academia and industry have joined forces and have contributed a number of systems allowing for the storage and processing of vast amounts of data over large numbers of computing nodes. This course will provide students with: (a) an understanding of the challenges introduced by Big Data, (b) an overview of data modelling and programming paradigms introduced by said systems, and (c) an overview of the design decisions underlying some of the most popular Big Data management systems (including Hadoop/MapReduce, Spark, and NoSQL data stores).		Good knowledge of database and operating systems concepts; basic understanding of computer networking and distributed systems concepts
BD025	Big data management systems for fun and profit	Dr Nikos Ntarmos, University of Glasgow	<b>28/07/2017</b> 09.00-17.00	Intermediate/ Advanced 1 day (6 hours)	The need to deal with ever increasing amounts of data has always led to the design and implementation of ground-breaking hardware and software systems ever since the first computers came to be. On the front of large-scale data management, the latest advances in a long line of innovative solutions have come in the	This is a highly technical course, so prior knowledge of Java and/or Python/Scala is an absolute	Attendance of the principles of big data storage and processing systems course BD024 (or similar), excellent



Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					form of highly scalable, fault-tolerant, reliable and efficient distributed systems. This course will focus on a select few such systems – namely HDFS, Hadoop/MapReduce, Spark, and Hbase – currently dominating the Big Data management landscape. By the end of this course, students will have a deep understanding of the design and implementation of these systems, will be able to explain the related trade-offs, fallacies and pitfalls, and will be able to write code that harnesses these systems to store and analyse large amounts of data.	requirement	knowledge of Java
BD026	Best practice in licensing and governance for research data - CANCELLED	UK Data Archive	<b>01/08/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	<p>In this hands-on day-long course, you will be introduced to best practices in licensing and governance models for research data. The day is run by data practitioners, thus the focus is on practical and efficient ways to get data assets for research through the pipeline from data creator to user. Ethical and legal issues will be covered so far as they apply to data, with real life examples from data collections available via the UK Data Service. We address a variety of data sources from academic data to bigger data sources from public and civil society bodies, and commercial organisations.</p> <p>The course will showcase the approach used by the UK Data Archive that enables this: our Data Access Policy and the Five Safes framework – enabling a data publisher to provide access to data across a spectrum, from open through safeguarded to controlled access. The philosophy we promote is to make data <b>open where possible and closed where necessary</b>.</p> <p>Participants will work in groups with hands on exercises including lightweight quizzes and scenario planning. We encourage participants to come prepared with a dataset in mind, so that</p>		A background in managing data assets in any organisation is useful, as is an awareness of data integrity and data protection. Technical experience is not needed

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					the experience of the day can be applied to this. This will be used in the final session to work up a plan for optimal data access that meets relevant ethical and legal restrictions		
BD027	Pictures – the power and the glory for great analytics	Dr Pete Barnsley, BT Openreach	<b>25/07/2017</b> 09.00-17.00	Introductory/ Intermediate, 1 day (6 hours)	<p>This session looks at the key first step in analytical studies and takes it through to developing data to create the insight. It will be a very interactive session using whiteboards and (if it can be arranged) SQL analysis of a datasets to show the linkage between the picture and the solution code.</p> <p>Here is the situation: you have a problem to work on - probably working in a project team with people from other disciplines (business improvement, project management) as well as sponsors and problem subject matter experts. A picture of the real world situation and all the aspects that could be affecting the situation clarifies the types of data needed. This could be classifiers (descriptors) or variables. The picture also acts as a way of gaining common understanding across the project team. The picture also helps communicate the problem and how it is being analysed.</p> <p>So how can we do this... come and pick up some tips and add your experience too</p>	Anyone interested in developing or moving into an analytical role. People working on developing data and answers to analytical problems	SQL capability and familiarisation of database data querying
BD028	Introduction to Machine Learning and Kernel Machines	Dr Luca Citi, University of Essex	<b>28/07/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	<p>The aim of this course is to provide an introduction to Machine Learning and a discussion of the types of problems it is suitable for. The course will then introduce Kernel Machines and show how they can provide robust but flexible classifiers when the number of training points is limited.</p> <p>Course objectives: - Understand how ML combines information from data and prior knowledge</p>	Background in computer science, maths, business and social science with some mathematical skills	The course requires high-school level maths. It will show the use of python for ML but students will have an option to code themselves or use the GUI-based software

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					<ul style="list-style-type: none"> <li>- Familiarise with supervised, unsupervised and reinforcement learning problems</li> <li>- Understand loss functions: the 0-1 loss for classification and its shortcomings</li> <li>- Understand how kernel classifiers based on the hinge loss can provide robust yet flexible classifiers</li> </ul>		Weka
BD029	Modelling and analysis of complex systems	Dr Chris Antonopoulos	<b>25/07/2017</b> 09.00-17.00	Introductory/ Intermediate 1 day (6 hours)	<p>This course is an introduction on the concepts and techniques of modelling and analysis developed in the emerging interdisciplinary field of complex systems. Complex systems can be realised as networks of many interacting components that arise and evolve through self-organisation. Many real-world systems can be modelled as complex systems, such as political organisations, human cultures and languages, national and international economies, stock markets, the Internet, social networks, the global climate, brains of living organisms, gene regulatory networks within cells, etc. The study of complex systems in a unified framework has been recognised in recent years as a new scientific discipline, the ultimate of interdisciplinary fields. In this course we will explore the universal, physical and mathematical principles that govern the emergence of complex systems from simple components, including the fundamentals of modelling, basics of dynamical systems, discrete- and continuous-time models, bifurcations, chaos theory, and finally static and dynamic networks.<sup>1</sup></p> <p><sup>1</sup> The material for this course is based on the book by H. Sayama, "Introduction to the</p>	Background in Mathematics, Physics, Computer Science, Social Sciences, Business	None

<sup>1</sup> The material for this course is based on the book by H. Sayama, "Introduction to the Modelling and Analysis of Complex Systems", OPEN SUNY Textbooks, Milne Library, 2015.

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					Modelling and Analysis of Complex Systems”, OPEN SUNY Textbooks, Milne Library, 2015.		
BD030	An introduction to Geographical Information Systems	Professor Andrew Lovett, University of East Anglia	<b>26/07/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	The course will provide an introduction to Geographical Information Systems, digital data sources and the QGIS open source software. It will consist of an introductory talk and a series of practical lab exercises (using the QGIS software) that will allow attendees to put their acquired skills into practice	Anyone interested in knowing more about GIS	No prior experience of GIS is required. Familiarity with Windows and Microsoft Office will be assumed
BD031	Best practice analytics	Professor Detlef Nauck British Telecom	<b>02/08/2017</b> 09.00-17.00	Intermediate/ advanced 1 day (6 hours)	Data Analytics has become a pillar of scientific research and decision support in businesses. With more data and more analytic tools becoming available to a larger user base across academia and businesses we are already seeing a “democratisation and commoditisation” of analytics. The proliferation of analytics makes it important to recipients of results to distinguish good analytics from bad. For analysts it means they should be able to demonstrate they followed best practice and that their results add value to research and business decision making. This course will look at methods and tools that can help us create high-quality analytics and reproducible results. We will also look at how to move from a single analyst, spreadsheet driven approach to collaborative analytics that follows a best practice governance model. Adopting practices from areas like software development, we will look at how to establish an analytics process based on documentation, versioning, testing, peer review, collaboration and risk evaluation. We will answer questions like ‘how can I make sure that my analysis is rigorous, complete and reproducible’, ‘how can I know the data I am using is any good’, ‘what steps should I follow to make sure the outcome of my analytics		We will use R for most examples, so attendance to course Introduction to R (BD009) or good working knowledge of R is advisable. A basic understanding of machine learning principles and some SQL will be useful but is not compulsory. <b>Participants are required to bring their own laptops</b> and install R ( <a href="http://www.r-project.org">www.r-project.org</a> ) , Rstudio ( <a href="http://www.rstudio.com">www.rstudio.com</a> ), Jupyter Notebooks ( <a href="http://jupyter.org">jupyter.org</a> ) and the free version

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					<p>project is reliable and usable’, ‘how can I collaborate in a group of analysts and share the work in a project’. We will use R, SQL and Python embedded in Jupyter Notebooks as an environment to work through a number of examples and use RapidMiner to demonstrate how to use a machine learning environment effectively.</p> <p><b>Course aim:</b> To give you an understanding of the challenges you will face when running your own real-world data analytics project and introduce you to a number of principles you can follow to achieve high-quality reproducible results.</p>		of RapidMiner ( <a href="http://rapidminer.com">rapidminer.com</a> ) beforehand.
BD032	Data science meets optimisation	Dr Andrew Parkes, University of Nottingham	<b>03/08/2017</b> 09.00-17.00	Intermediate, 1 day (6 hours)	<p>A wide variety of optimisation techniques have been developed for many real-world decision making problems such as scheduling, timetabling, vehicle routing, and many others. These techniques themselves generally involve many different heuristic choices, and usually as part of a long sequence of internal decisions. A simple example would be the sequence of temperatures used within simulated annealing. Increasingly it is being recognised that the control of such sequences of internal decisions would benefit from application of data science techniques. This gives an opportunity for data science to improve existing optimisation methods in many real world problems. Conversely, data science itself can often be framed as optimisation problem. Hence there are potentially opportunities for a useful interplay between the two sides, and the course will provide foundations for this. That is, illustrating how data science may be applied to traditional optimisation, and complemented with illustrations of how modern optimisation relates to data science</p>		Some knowledge of optimisation problems and techniques is useful but not essential

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
BD033	Introduction to Artificial Neural Networks and Deep Learning with TensorFlow	Dr Michael Fairbank, University of Essex	<b>27/07/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	This tutorial will make an introduction to neural networks and deep learning. The course is aimed at general scientists with some familiarity of computer programming, but who are new to deep learning. The tutorial will use TensorFlow and Python as the implementation languages. It will start by covering simple feed-forward neural networks, activation functions, loss functions, backpropagation and gradient descent theory, illustrated by practical exercises. It will then describe and implement deeper neural network architectures such as convolutional and recurrent neural networks. Simple trial applications from image processing and time sequence analysis will be used as examples. By the end of the tutorial, all participants should have achieved a basic understanding of deep learning, and have made several successful basic implementations using TensorFlow		Familiarity with any computer programming language (but preferably Python), and familiarity with matrix multiplication
BD034	Science and data	Dr Andrew Harrison, University of Essex	<b>31/07/2017</b> 09.00-17.00	Introductory, 1 day (6 hours)	The scientific world view has arisen from integrative analyses of data sources. I will discuss three “big picture” historical examples, from physics, biology and geology. I will present an overview of the current data explosion and the opportunities that this offers for new ways of working. But I will also note caution, due to both the reproducibility crises that is now acting to discredit many fields of research and the spurious correlations seen within large data sets. I will also describe the growth of the open movement.		None
BD035	<b>How can Autonomous Cars see and understand their environment</b>	Dr Aura Hernandez Sabate, Universitat Autònoma de Barcelona	<b>25/07/2017</b> 09.00-17.00	6 hours	Autonomous cars need to manage large amounts of data to see and understand their environment and perform in harmony with what is happening around them. Computer Vision and machine learning are powerful tools to help them in these tasks. With this course, we will take a tour along		

Course code and title		Presenter	Date and Time	Level and Duration	Abstract	Target audience	Prerequisites/ Requirements
					the advances of computer vision and machine learning over the years focusing on problems related to the autonomous performance of cars. In particular, we will pose the problem of pedestrian detection and scene segmentation to show the evolution of the solutions proposed by the Computer Vision community		

### Keynote Lectures

Ref	Title	Presenter	Date/time	Abstract
<b>KN001</b>	Have your own artificially intelligent Space Program	John McNamara, Senior Inventor, IBM	<b>24/07/2017</b> 17.15-18.00	How you can build and run your own space program, using Watson, Cloud and IoT for less than £500. How do to it. And how not to do it
<b>KN002</b>	Reshaping public services, increasing prosperity, improving lives: Digital at the heart of whole system change for Essex	David Wilde, Director of Digital, Essex County Council	<b>26/07/2017</b> 17.30-18.30	There is no question that digital is fundamentally changing the way we live, work and spend our leisure time. It has also bulldozed its way through many market sectors, disrupting the order of things in months instead of years or decades. In Essex we want to embrace the potential rather than hold the line on traditional practice, which means looking at what is possible through the eyes of residents and businesses rather than the institutions we have today. Our digital strategy, driven by our Smart Essex ambition, sets out our approach to turning this into reality, recognising that the change needed has to be driven from within
<b>KN003</b>	Analysis and Forecasting of Complex Dynamical Systems with Neural Networks	Hans-Georg Zimmerman, Senior Principal Research Scientist, Siemens Corporate Technology, Germany	<b>31/07/2017</b> 17.15-18.00	Neural networks have been proven to be universal approximators but this still leaves the identification task a hard one. Beside using data we should focus our attention on the underlying structure of our subject of interest. In case of dynamical systems this is time, leading us to state space models and recurrent neural networks. After an introduction of small (open) dynamical systems we will study dynamical systems on manifolds. Here manifold and dynamics have to be identified in parallel. We will move on to large (closed) dynamical systems with hundreds of state variables and will combine causal and retro-causal models of the observations. This combination leads us to an implicit description of dynamical systems on manifolds. Finally we will discuss the quantification of uncertainty in forecasting. In our framework the uncertainty appears as a consequence of principally unidentifiable hidden variables in the description of large systems. Together with the mathematical concepts we will see applications in economics and

				engineering
<b>KN004</b>	Beyond the Data and Digital transformation: Becoming a smart organisation.	Antonio Alvarez, Head of Data Innovation, Isban UK, Santander Group	<b>02/08/2017</b> 17.15-18.00	In the current world of ever faster pace of change and increasingly disrupted industries, Santander UK is undertaking a metamorphosis to become a data driven digital organisation. We will discuss some of the lessons learned during this journey and how they have created the consciousness that we need to aim beyond digital, into the adoption of a scientific mind set throughout the company with the application of advance analytics and machine learning