Course Name	**MSc in Digital Scholarship** # Machine Learning and AI
Course Leader	**Christine Sheldon** (Politics/Data Science) c.r.sheldon@bham.ac.uk
Introduction	This course aims to equip students with the knowledge and skills necessary to employ Machine Learning (ML) methods in your own research, serving as a stepping stone for continued learning and development in this exciting area. Throughout this course, we will focus on the practical aspects of machine learning, without requiring significant background knowledge of the underlying algebraic foundations. This approach will enable you to quickly grasp essential concepts and techniques, allowing you to apply them effectively in your research projects. This will be done using the Python and the Scikit-Learn package. As a result, basic prior knowledge of Python is required. However, no prior experience in machine learning using Python is necessary, as you will be guided through the fundamental concepts and tools needed to get started. By the end of this course, you will be familiar with a wide range of machine learning techniques, including supervised learning, unsupervised learning, and natural language processing. You will also have the opportunity to work on your own project, giving you hands-on experience in applying machine learning.
Reading List	Books used in the course (Exact chapters and sections listed under the relevant classes): ● Critical Digital Humanities : The Search for a Methodology by James E. Dobson. ● Machine Learners: Archaeology of a Data Practice by Adrian Mackenzie. ● Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition, by Aurélien Géron. ● Introduction to Machine Learning with Python by Andreas C. Müller, and Sarah Guido. ● Python Machine Learning - Third Edition by Sebastian Raschka, Vahid Mirjalili ● Practical Python Data Visualization: A Fast Track Approach To Learning Data Visualization With Python ● Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. ● Artificial intelligence and its discontents: critiques from the social sciences and humanities edited by Ariane Hanemaayer

- Natural Language Processing with Python by Steven Bird, Ewan Klein, and Edward Loper
- Text as Data by Justin Grimmer, Margaret Roberts, and Brandon Stewart.
- Generative Deep Learning, 2nd Edition, by David Foster

The majority of books are available via the O'Reilly platform. The following link leads to a curated playlist, including videos:
https://learning.oreilly.com/playlists/7866bbd5-0353-49dc-a4e8-655161b07034

Week 1	Introduction to Machine Learning

1.1   Introduction to Machine Learning and This Course

Key Reading
- Chapter one from Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition, by Aurélien Géron. Available on the O'Reilly platform.

Videos
- Lesson 1.1.1 and 1.1.2 from Machine Learning in Python for Everyone (Video Collection) by Mark Fenner. Available on the O'Reilly platform.

Visualising Articles
- How Machines Learn:
  https://www.ft.com/content/6692ba73-fa71-45db-881c-0be221495b3a

Supplementary Reading
- Chapter one from Machine Learners: Archaeology of a Data Practice by Adrian Mackenzie. Available on SOLO.
- Chapter two from Critical Digital Humanities : The Search for a Methodology by James E. Dobson. Available on SOLO.

1.2   Data Collection and Preparation for ML Applications

Key Reading
- Chapter one *until 1.7*, from Introduction to Machine Learning with Python by Andreas C. Müller, and Sarah Guido. Available on the O'Reilly platform.
- Chapter 4 *until 'Selecting Meaningful Features'* of Python Machine Learning - Third Edition by Sebastian Raschka, Vahid Mirjalili. Available on the O'Reilly platform.

Supplementary Reading
- Web Scraping with Python, 3rd Edition, by Ryan Mitchell. Available on the O'Reilly platform.

1.3   Exploratory Data Analysis and Visualization

Key Reading
- Chapter 5 from Practical Python Data Visualization: A Fast Track Approach To Learning Data Visualization With Python by Ashwin Pajankar. Available on the O'Reilly platform.

Videos
- 2.8 and 2.9 of Data Science and Machine Learning with Python – Hands-On! by Frank Kane. Available on the O'Reilly platform.

Supplementary Reading
- Fundamentals of Data Visualization by Claus O. Wilke. Available on the O'Reilly platform.

Week 2	Supervised Learning

2.1   Regression algorithms

Key Reading:
- Chapter 3: Linear Regression, until 3.2.2, from An Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Available on SOLO.

Videos:
- Linear Regression and Multiple Regression from Chapter 3 of Data Science and Machine Learning with Python - Hands-On! by Frank Kane. Available on the O'Reilly platform.

2.2   Classification algorithms

Key Readings:
- Chapter Two *until 2.4* from Introduction to Machine Learning with Python by Andreas C. Müller, Sarah Guido. Available on the O'Reilly platform.
- Chapter 4 *until 4.4* from An Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Available on SOLO

Videos
- Lesson 3.2.1 and 3.3.3 from Machine Learning in Python for Everyone (Video Collection) by Mark Fenner. Available on the O'Reilly platform.

Visualisation Articles
- Decision Trees: http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

2.3   Model optimization and Performance Metrics.

> **Key Readings:**
> - Chapter 5 from Introduction to Machine Learning with Python and Andreas C. Müller, Sarah Guido. Available on the O'Reilly platform.
>
> **Videos:**
> - 4.1 and 4.2 of Data Science and Machine Learning with Python – Hands-On! by Frank Kane. Available on the O'Reilly platform.
> - Lesson 2.2.1 -2.2.4 from Machine Learning in Python for Everyone (Video Collection) by Mark Fenner. Available on the O'Reilly platform.
>
> **Visualisation Articles**
> - Model Tuning & Bias-Variance Trade Off: http://www.r2d3.us/visual-intro-to-machine-learning-part-2/

Week 3	Unsupervised Learning

3.1   Dimensionality Reduction - Principal Component Analysis (PCA)

> **Key Readings:**
> - Chapter 12 *until 12.4* from an Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Available on SOLO
> - Chapter 8 *until 'Random Projection'* from Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition, by Aurélien Géron. Available on the O'Reilly platform.
>
> **Videos:**
> - 6.3 and 6.4 of Data Science and Machine Learning with Python – Hands-On! by Frank Kane. Available on the O'Reilly platform.
>
> **Visualisation Articles**
> - Understanding UMAP: https://pair-code.github.io/understanding-umap/

3.2   Clustering algorithms - K-means

> **Key Readings:**
> - Chapter 12.4 from an Introduction to Statistical Learning by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. Available on SOLO

- Chapter 11 from Python Machine Learning - Third Edition by Sebastian Raschka and Vahid Mirjalili. Available on the O'Reilly platform.

Supplementary Reading:
- Chapter 12 and 13 of Text as Data by Justin Grimmer, Margaret Roberts, and Brandon Stewart. Available on SOLO.

3.3   Ethics and Fairness in ML

Key Readings:
- Artificial Intelligence: A Medium that Hides Its Nature by Avon Huxor. A chapter from Artificial intelligence and its discontents: critiques from the social sciences and humanities edited by Ariane Hanemaayer. Available on SOLO.
- Gender Bias in Machine Translation Systems by Stephanie Ullmann. A chapter from Artificial intelligence and its discontents: critiques from the social sciences and humanities edited by Ariane Hanemaayer. Available on SOLO.

Week 4	**Natural Language Processing (NLP)**

4.1   Text processing: Tokenization, Stemming, and Lemmatization.

Key Readings:
- Chapter 5 from Text as Data by Justin Grimmer, Margaret Roberts, and Brandon Stewart.

Supplementary Readings:
- Chapter 3 from *Regular Expressions for Detecting Word Patterns* onwards, from Natural Language Processing with Python by Steven Bird, Ewan Klein, and Edward Loper

4.2   NLP applications - Sentiment Analysis, Topic Modelling, and Text Classification.

Key Readings:
- Chapter 7 from Introduction to Machine Learning with Python by Andreas C. Müller and Sarah Guido. Available on the O'Reilly platform.

4.3   Word Embeddings and Large Language Models (LLM)

Key Readings:
- Chapter 8 from Text as Data by Justin Grimmer, Margaret Roberts, and Brandon Stewart.
- Wankmüller. (2022). Introduction to Neural Transfer Learning With Transformers for Social Science Text Analysis. Sociological Methods & Research,4912412211345–.https://doi.org/10.1177/00491241221134527

Visualising Articles

- Generative AI https://ig.ft.com/generative-ai/

Supplementary Reading
- Chapter 9 from Generative Deep Learning, 2nd Edition, by David Foster

Assessment	For the final assignment students can either write a ML algorithm using data of their own choosing; or replicate (part of) an academic article using ML. The annotated code and an accompanying report will be due at the end of the course, and comprise a 100% of the grade.
Learning Outcomes	Upon completing this course, students will: ● Gain a foundational understanding of machine learning concepts and techniques, including supervised learning, unsupervised learning, and natural language processing. ● Develop practical skills in implementing machine learning algorithms using Python and the Scikit-Learn package. ● Learn how to pre-process data, select appropriate machine learning models, and evaluate their performance. ● Understand the ethical implications of using machine learning in research, including potential biases and privacy concerns. ● Build the confidence to continue learning and exploring more advanced machine learning techniques and applications in the digital humanities field.
Course Leader Information	Christine Sheldon is a Research Fellow in Computational Social Science at the University of Birmingham. Previously she completed her doctorate at the Department of politics and International Relations (DPIR) of the University of Oxford. She teaches courses on Social Data Science and Machine Learning at a number of different departments of the University of Oxford. In her own research she deploys ML and NLP techniques to measure and analyse latent variables in political text, and designs AI systems for public service provision.