2024 / 25

School of Science and Computing
+353 (0)51 302037
Eleanor.Reade@setu.ie
www.wit.ie/schools/science_computing

Ollscoil Teicneolaíochta an Oirdheiscirt

South East Technological University

Module Descriptor

Building Data Science Models (Computing and Mathematics)

Building Data Science Models (A33701)

Short Title: Building Data Science Models
Department: Computing and Mathematics
Credits: 10 **Level:** Advanced

Description of Module / Aims

This module introduces the student to the practice of data science, where techniques and algorithms from mathematics and statistics, supplemented with advanced data infrastructures and processes, are used to learn from data. In this context, learning includes understanding and prediction, and the data might not have been collected for the purpose of such analysis. The student will be introduced to the machine learning pipeline, which provides a principled, mathematically rigorous way to build data science models. With these models, and the solution algorithms covered in the module, the student can extract understanding and knowledge from the data. The practical part of the module will present a suite of machine learning exercises that the student will complete by building data science models, making predictions, validating those predictions, and interpreting the results.

Programmes

	stage/semester/status
HDip in Science in Data Analytics (WD_KDAAN_G)	2 / 4 / M

Indicative Content

- Exploratory Data Analysis: preparing, visualising, structuring and understanding data.
- Data cleaning and pre-processing.
- Selecting models by type: regression, classification, clustering, etc.
- Clustering: hierarchical, k-means, k-center, role of distance metrics.
- Classification: Naive-Bayes, support vector machines, logistic regression, decision trees.
- Ensembles: random forest, etc., bagging and boosting methods (e.g., XGBoost, AdaBoost, CART aggregation etc.), stacking models.
- Model evaluation and optimisation: Feature extraction/selection/engineering, model selection, hyperparameter tuning using grid/random and Bayesian optimisation.
- Presenting the results and indicating their limitations and assumptions.
- Explainable AI: black vs glass box models, local vs global models (LIME, SHAP, etc.).
- Using pre-trained models: transfer learning and fine-tuning.
- Model deployment: Review of current frameworks and tooling, model drift.

Learning Outcomes

On successful completion of this module, a student will be able to:

1. Categorise typical fundamental Data Science problems.
2. Appraise the concepts and fundamentals of core machine learning techniques such as Regression, Classification and Clustering and the algorithms that are used for each.
3. Evaluate the use of typical data science techniques and their appropriate implementation.
4. Implement an appropriate ML model and evaluate its performance for given datasets.
5. Solve data science problems using a rigorous and repeatable machine learning pipeline.
6. Assemble and interpret the results of a data science-based study.
7. Deploy and monitor a ML model.

Learning and Teaching Methods

- The lectures will introduce data science theory and practice to the student. The student will be encouraged to participate in class discussions and ask questions to support their learning process.
- The practical classes extend the lectures by showing data science in action and encouraging the student to use and extend the concepts.
- Students will apply typical machine learning methods to data sets provided in the practical classes.
- Students will interpret and present, in context, the findings produced in the practical classes and continuous assessment.

Learning Modes

Learning Type	F/T Hours	P/T Hours
Lecture	48	48
Practical	48	48
Independent Learning	174	174

Assessment Methods

	Weighting	Outcomes Assessed
Continuous Assessment	100%	
Assignment	15%	1,2
Assignment	25%	1,2,3,4,6
Assignment	25%	2,3
Assignment	35%	2,4,5,6,7

Assessment Criteria

<40%: Unable to interpret and describe key concepts of data science.

40%–49%: Be able to interpret and describe key concepts of data science.

50%–59%: Ability to discuss and use key concepts of data science and ability to discover and integrate related knowledge in other knowledge domains.

60%–69%: Be able to solve problems using data science techniques by applying the appropriate skills and tools.

70%–100%: All the above to an excellent level. Be able to analyse and design solutions to a high standard for a range of both complex and unforeseen problems through the use and modification of appropriate skills and tools.

Supplementary Material(s)

- Geron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow.* CA: O'Reilly, 2019.
- JiaweiHan, H., M. Kamber and J. Pei. *Data Mining, Third Edition: Concepts and Techniques.* 3rd. NY: Morgan Kaufmann, 2011.
- Pakshaver, B. *Pandas in Action.* NY: Manning, 2021.