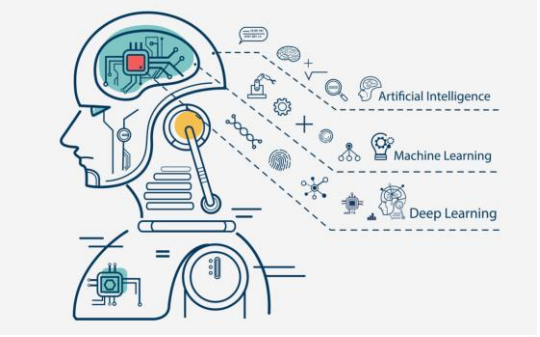

cct College Dublin
Computing • IT • Business

Machine Learning for Data Analysis

MSc in Data Analytics

Module Introduction

CCT College Dublin

Ireland

Introduction

cct | College Dublin
Computing • IT • Business

- **Lecturer:** Dr. Muhammad Iqbal*
- **Experience:** Data Analytics and Processing, Numerical Modelling & Simulations, Structured & Object-Oriented Programming, Data Structures & Algorithms, Scalable Systems Programming (Python, R, Matlab, etc..).
- **E-mail:** miqbal@cct.ie
- **Contact:** Use CCT email address for contact along with your **Module and Course names**.

Module Information

cct | College Dublin
Computing • IT • Business

- **Contact hours:**
  - 2.5 hours lecture and tutorial
  - More than 5 – 6 hours weekly independent learning
- **Continuous Assessments**
  - 100% Continuous Integrated Assessment
- **Machine Requirements**
  - Windows 10/ 11 machine

Objectives

cct | College Dublin
Computing • IT • Business

- The underlying concepts of machine learning are mentioned below

1. The different categories of machine learning techniques.
2. The different stages of the Knowledge Discovery life cycle.
3. The major Supervised, Unsupervised and Semi-Supervised learning techniques.
4. The application, optimisation and validation of various machine learning techniques.

Learning Outcomes

cct | College Dublin
Computing • IT • Business

- **On successful completion of this module, the learner will be able to**

1. Modify and implement Machine Learning Algorithms to solve analytical problems. (Linked to PLO 1, PLO 2, PLO 5)

2. Determine whether a given data analysis problem requires the use of **supervised, semi-supervised or unsupervised learning** methods. Develop and implement the chosen learning method. (Linked to PLO 1, PLO 2, PLO 4)

3. Develop a machine learning strategy for a given domain and communicate effectively to team members, peers and project stakeholders the insight to be gained from the interpreted results. (Linked to PLO 1, PLO 4, PLO 6)

4. Implement a range of **classification and regression techniques** and detail/ document their suitability for a variety of problem domains. (Linked to PLO 5)

5. Critically evaluate the performance of Machine Learning models, propose strategies to optimise performance. (Linked to PLO 3)

Topics

Content
Syllabus rationale : Implementation of theoretical concepts Introduction • Supervised, semi-supervised and unsupervised learning (conceptual) • Classification and Regression (integrated concept : Statistics for Data Analysis) (conceptual) • Machine Learning, Deep Learning and Reinforcement Learning (conceptual) • CRISP-DM, KDD and SEMMA (practical)
Supervised Learning (integrated concept : Statistics for Data Analysis / Data Preparation and Visualisation) • Linear Regression (practical) • Nearest Neighbour (practical) • Gaussian Naive Bayes (practical) • Decision Trees (practical) • Support Vector Machine (SVM) (practical) • Random Forest (practical)
Unsupervised Learning (integrated concept : Data Preparation and Visualisation) • Clustering (practical) • Association (practical) • Anomaly Detection (practical) • Dimensionality Reduction (practical)
Semi-Supervised Learning (integrated concept : Data Preparation and Visualisation) • Natural Language Processing (practical)
Case Studies (practical examples) • Supervised Learning • Unsupervised Learning • Semi-Supervised Learning • Reinforcement Learning • Deep Learning
Validation and Optimisation (integrated concept : Statistics for Data Analysis) • Validation (Re-substitution, Hold-out, K-fold cross-validation, LOOCV, Random subsampling, Bootstrapping...) (practical) • Optimisation (loss functions/cost functions, Gradient Descent, Momentum, AdaGrad, RMSProp, Adam...) (practical)

cct College Dublin
Computing • IT • Business

CCT Resources

cct | College Dublin
Computing • IT • Business

- Essentials 1- Study Skills (https://moodle.cct.ie/course/view.php?id=677)
- Essentials 2 - Research Skills (https://moodle.cct.ie/course/view.php?id=678)
- CCT ARC (https://arc.cct.ie/)
- https://www.datacamp.com/community
- www.datacamp.com
- For technical support, contact with Mr. Juan Murguey

cct | College Dublin
Computing • IT • Business

Questions?

Books and eBooks

cct College Dublin
Computing • IT • Business

- Aurélien Géron, 2019, 2nd Edition, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reilly Media [ISBN: 978-1492032649]
- Andriy Burkov, 2019, The Hundred-Page Machine Learning Book, Andriy Burkov, [ISBN: 978-1999579500]
- Introduction to Machine Learning with Python, Andreas C. Müller and Sarah Guido, O'Reilly Media, Inc. October 2016.
- Thakur, A. (2020) Approaching (Almost) Any Machine Learning Problem. London: Abhishek Thakur. ISBN: 9788269211504.
- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, Aurélien Géron, O'Reilly Media, September 2019, ISBN: 9781492032649.
- Discovering Knowledge In Data: An Introduction To Data Exploration, Second Edition, By Daniel Larose And Chantal Larose, John Wiley And Sons, Inc., 2014.