MACHINE LEARNING & ANALYTICS FOR ECONOMISTS

IE University
Professor: **ALBERTO TORRES BARRÁN**
E-mail: atorresb@faculty.ie.edu

Academic year: 23-24
Degree course: FOURTH
Semester: 2º
Category: COMPULSORY
Number of credits: 3.0
Language: English

PREREQUISITES

Basic statistical and mathematical concepts

SUBJECT DESCRIPTION

Machine Learning poses both opportunities and challenges for businesses. No matter the industry or role in which you are planning on working, and no matter the type of organization – as a professional, your world will be awash with data. The most successful managers today are those who know how to drive innovation and performance improvement using data. You need to be able to understand and be comfortable with big data terminology, analytical methods and be able to work and make decisions using quantitative information.

This course will help you develop the analytical skills you need now in order to summarize data, find meaning of it, and extract its value to drive decisions and change. In this course you will be exposed to numerous real-world examples of how Machine Learning is reshaping functional areas as well as innovation and entrepreneurship.

This is a course for students who are interested in adding Machine Learning and analytical thinking to their general management or who are interested in working directly or starting a company related to Machine Learning.

OBJECTIVES AND SKILLS

The main objective of this first course is to learn how leading companies in a range of industries are using Machine Learning and Data Analytics. For that, you will be introduced in the essential practice of some machine learning algorithms commonly used for data science applications and predictive analytics.
The course is not only about reviewing some real world experiences but about being able to run basic analysis and to clearly understand the use and interpretation of a collection of algorithms commonly used in real data science applications. The amount of mathematics and statistical concepts will be reduced in this course, expending most of the time in:

- Discovering Machine Learning potential with a strategic view
- Acquiring a solid understanding of the algorithms and tools used by data scientists emphasizing

the applications of machine learning algorithms in every day's practice in real business.

- Practicing, at an elementary level, how to run simple analysis and properly interpret and present the results obtained for helping decision making processes.

The expected outcomes of the course are:

- To acquire an integrated approach to data as a strategic asset and to understand the real usefulness of data analytics techniques in a real world.
- To have a broad panorama of the different business contexts in which different machine learning algorithms become a valuable resource.
- To learn the motivation behind every different technique, relative advantages and disadvantages, when to apply each one and how to come to meaningful conclusions.
- To gain a practical understanding of how to fruitfully conduct machine learning oriented applications by using data science tools at the elementary level.
- To avoid common mistakes when using or interpreting modeling outcomes.

METHODOLOGY

The course is mainly a practical hands-on course. During the sessions we will mix theoretical lecturing with practical assignments. The course is designed for you to always have a laptop with you in class so you can code directly the exercises during the sessions. Approximately 50% of the time in the sessions we will have class discussions and theoretical lecturing together with small exercises to be completed during the session.

- **Module 1** (Sessions 1 to 3) is organized around presentation of concepts and active discussions on business cases. It is important that you come to class prepared to enrich class discussions.
- **Module 2** (Sessions 4 to 9) combines synthetic datasets and a cloud-based data science tool to introduce key machine learning concepts. Classes will be based on practical activities with the tool through targeted examples.
- **Module 3** (Sessions 10 to 12) is devoted to understand more complex machine learning applications using the concepts gained in the previous modules. Likewise, it is important that you come to class prepared to enrich class discussions.
- **Module 4** (Sessions 13 and 15) aims to evaluate the acquisition of the concepts and techniques by the students via a group assignment and an exam.

Teaching methodology	**Weighting**	**Estimated time a student should dedicate to prepare for and participate in**
Lectures	21.33 %	16 hours
Discussions	6.67 %	5 hours
Exercises	18.67 %	14 hours
Group work	26.67 %	20 hours
Other individual studying	26.67 %	20 hours
TOTAL	100.0 %	75 hours

PROGRAM

SESSION 1 (LIVE IN-PERSON)

Introduction to Machine Learning

Book Chapters: 50 years of Data Science (Presentation at the Tukey Centennial workshop, Princeton, NJ, Sept 18 2015) (CED)
Article: A Few Useful Things to Know About Machine Learning (Communications of the ACM, October 2012, Vol. 55, No. 10) (CED)
Other / Complementary Documentation: What is the difference between data mining, statistics, machine learning and AI? (StackExchange.com)

- Syllabus, objectives, content organization, tools, rules and grading.
- What can Machine Learning do and not do?
- Machine Learning applications
- Machine Learning core concepts

SESSION 2 (LIVE IN-PERSON)

Machine Learning pipelines

- Building an Machine Learning project
- Machine Learning lifecycle
- More Machine Learning examples
- Problem types: classification vs regression

SESSION 3 (LIVE IN-PERSON)

Machine Learning tools: BigML

Video: BigML Education Videos (BigML)

- ML tools landscape
- Task: Create free BigML account

SESSION 4 (LIVE IN-PERSON)

Classification

- K-nearest neighbours
- Logistic Regression
- Overfitting and hyper-parameter tuning

SESSION 5 (LIVE IN-PERSON)

Linear models for regression

- Linear regression
- Regularization
- Linear regression variants

SESSION 6 (LIVE IN-PERSON)

Evaluating Machine Learning models

Book Chapters: Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning (arXiv:1811.12808v3, Cornell University) (CED)

- Classification and regression metrics
- Train-test-validation splits
- Cross validation

SESSION 7 (LIVE IN-PERSON)

Decision trees and ensemble models

- Decision trees
- Ensemble models: bagging and boosting
- Random Forest
- Gradient boosting

SESSION 8 (LIVE IN-PERSON)

Unsupervised learning

- Introduction to non-supervised learning
- Clustering
    - K-means
- Anomaly detection
    - Isolation Forest

SESSION 9 (LIVE IN-PERSON)

Explainability and fairness

Article: Principles and Practice of Explainable Machine Learning (arXiv:2009.11698v1, Cornell University) (CED)
Article: Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges (arXiv:2010.09337v1, Cornell University) (CED)
Book Chapters: Interpretable Machine Learning (Independently published (28 febrero 2022))
Article: A Survey on Bias and Fairness in Machine Learning (arXiv:1908.09635v2, Cornell University) (CED)

- Bias in Machine Learning
- Common sources of bias
- Evaluation model fairness
- Explainability and interpretability of Machine Learning algorithms

SESSION 10 (LIVE IN-PERSON)

Neural Networks and Deep Learning

- Introduction to Neural Networks
- Perceptron
- Representation learning
- Feed-forward neural network

SESSION 11 (LIVE IN-PERSON)

Computer Vision

- Convolutional Neural Networks
- Working with images

SESSION 12 (LIVE IN-PERSON)

Natural Language Processing

Article: How to solve 90% of NLP problems: a step-by-step guide (Medium, Jan 24, 2018)

- Working with text data
- Bag of words
- Word embeddings

SESSION 13 (LIVE IN-PERSON)

Group presentations

Students will present their proposed Machine Learning project (See guidelines in section Evaluation Method).

SESSION 14 (LIVE IN-PERSON)

Revision session

Review of the main concepts and techniques learned over the course in order to prepare the final exam.

SESSION 15 (LIVE IN-PERSON)

Final exam

BIBLIOGRAPHY

Compulsory

- Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. (2022). *An Introduction to Statistical Learning.* 2nd. Springer. ISBN 9781071614204 (Printed)

Recommended

- Ian Goodfellow and Yoshua Bengio and Aaron Courville. *Deep Learning.* MIT Press. ISBN 0262035618 (Digital)

https://www.deeplearningbook.org/

EVALUATION CRITERIA

Criteria	Percentage	Comments
Final Exam	50 %	
Workgroups	20 %	
Group Presentation	20 %	
Class Participation	10 %	

The **Final Exam** will be graded between 0-100. Students must score at least 40 points in order to pass the course.

The **Class Participation** (10%) takes into consideration (i) your attendance, (ii) your participation in class and (iii) the quality of your participation.

The workgroup project is composed of a **Workgroup Written Report** (20%) and a **Group Presentation** (20%) (Session 13). Students will prepare a Written Report and Group Presentation on their proposed Machine Learning project as if they were presenting it to their manager. The presentation has to include details on both the rationale for the initiative and the implementation.

WRITTEN REPORT GUIDELINES

Your written report should not be more than 5 pages long, excluding 1 page Cover Sheet and Appendix (if applicable). Please use the following outline for the content of your report:

1. Cover Sheet (1 page) showing the full name of all authors, sector applied and, if applicable, a project name.
2. Existing Situation (~1 page). Description of the sector company to be analysed, current business model and competitors.
3. Proposed Solution (~ 1 page). Description of the rationale for the initiative and the proposed solution.
4. Data (~1 page). Description of the data expected to gather for your project.
5. Data Analysis (~1 page). Description of the proposed data analysis.
6. Expected Outcomes (~1 page). Competitive advantage thanks to the implementation of the proposed Big Data initiative in the company

The font used for your report can have a maximum size of 11 points.The grading criteria for the Workgroup Written Report will be: (i) originality (40%), (ii) soundness (40%) and (iii) clarity (20%).

GROUP PRESENTATION GUIDELINES

The group presentation consists of a presentation to the class of their proposed Machine Learning project (Session 13). The presentation is up to 10 minutes long, followed by questions. All students are encouraged to ask questions about other projects and any member of the group should be prepared to respond to such questions. Final grade of the project will be the same for all members of the project, independently of who has done the presentation. However, all members must contribute equally to the full project. Otherwise, the members that did not make a meaningful contribution will receive a grade of 0.

Please, plan your time carefully, and most importantly, prioritize the information you present. Moreover, the time constraint should encourage you to find the best way to communicate your insights in a convincing, clear and concise manner. Put yourself in a situation as if you were presenting to your manager. Both the content and the quality of the presenting skills will be evaluated, so practice and rehearse to ensure smooth flow, clarity of speech and enthusiastic attitude. Remember to include and initial slide with the project name and authors (full names of all team members).

Your presentation will be evaluated according to the following criteria: (i) demonstrated understanding of the machine learning project proposed (25%), (ii) quality and clarity of the presentation (25%), (iii) quality of the supporting material (25%), and (iv) overall interest that the presentation creates (25%).

RE-SIT / RE-TAKE POLICY

Each student has four (4) chances to pass any given course distributed over two (2) consecutive academic years. Each academic year consists of two calls: one (1) ordinary call (during the semester when the course is taking place); and one (1) extraordinary call (or “re-sit”) in June/July.

Students who do not comply with the 70% attendance requirement in each subject during the semester will automatically fail both calls (ordinary and extraordinary) for that Academic Year and have to re-take the course (i.e., re-enroll) during the next Academic Year.

The Extraordinary Call Evaluation criteria will be subject to the following rules:

- Students failing the course in the ordinary call (during the semester) will have to re-sit

evaluation for the course in June / July (except those students who do not comply with the attendance rule, and therefore will not have that opportunity, since they will fail both calls and must directly re-enroll in the course during the next Academic Year).

- It is not permitted to change the format nor the date of the extraordinary call exams or deadlines under any circumstance. All extraordinary call evaluation dates will be announced in advance and must be taken into consideration before planning the summer (e.g. internships, trips, holidays, etc.)
- The June/July re-sit will consist of a comprehensive evaluation of the course. Your final grade for the course will depend on the performance in this exam or evaluation only. I.e., continuous evaluation over the semester (e.g. participation, quizzes, projects and/or other grade components over the semester) will not be taken into consideration on the extraordinary call. Students will have to achieve the minimum passing grade of 5 and the maximum grade will be capped at 8.0 (out of 10.0) – i.e., "notable" in the extraordinary call.
- Re-takers: Students who failed the subject on a previous Academic Year and are now re-enrolled as re-takers in a course will need to check the syllabus of the assigned professor, as well as contact the professor individually, regarding the specific evaluation criteria for them as re-takers in the course during that semester (ordinary call of that Academic Year). The maximum grade that may be obtained as a retaker during the ordinary call (i.e., the 3rd call) is 10.0 (out of 10.0).

After exams and other assessments are graded by the professor (on either the ordinary or extraordinary call), students will have a possibility to attend a review session (whether it be a final exam, a final project, or the final overall grade in a given course). Please be available to attend the session in order to clarify any concerns you might have regarding your grade. Your professor will inform you about the time and place of the review session.

- Students failing more than 18 ECTS credits after the June/July re-sits will be asked to leave the Program. Please, make sure to prepare yourself well for the exams in order to pass your failed subjects.
- In case you decide to skip the opportunity to re-sit for an exam or evaluation during the June/July extraordinary call, you will need to enroll in that course again for the next Academic Year as a re-taker, and pay the corresponding tuition fees. As you know, students have a total of four (4) allowed calls to pass a given subject or course, in order to remain in the program.

PROFESSOR BIO

Professor: **ALBERTO TORRES BARRÁN**

E-mail: atorresb@faculty.ie.edu

Professor ALBERTO TORRES BARRÁN

Alberto Torres Barrán is a PhD in Computer Science by Universidad Autónoma de Madrid. Currently he is working as a Machine Learning Engineer at Komorebi.

More information:

- Twitter
- Github
- Linkdn

OTHER INFORMATION

OFFICE HOURS - CONTACT INFORMATION

- Office hours: Live tutorials available by previous appointment.
- Contact details: e-mail: atorresb@faculty.ie.edu