**Program: MBA**

**Specialization: Digital Transformation & Data Science**

**Syllabus for Module: International Immersion (Selected Topics in Digital Transformation & Data Science)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date:**  July 15-16, 2024 | **Lecturer:**  Ted Kwartler | **ECTS:**  2,5 | **Language:**  English |

**Course Description**

|  |
| --- |
| This intensive seminar on Data Science and Business Analytics, targeted towards Executive MBA students at WU EA, will delve into the analysis of diabetes patient readmission data—a prominent issue in healthcare management. The course will shepherd learners through the comprehensive data science workflow using R; commencing with problem understanding, continuing through the stages of obtaining data as well as cleaning & munging, exploratory data analysis (EDA), model development, and ultimately, model testing. By the end of this seminar, attendees will have acquired a robust understanding of leveraging data to drive insights in healthcare management, thus bridging the gap between data science and strategic decision-making in the healthcare sector.  Course material will be on our class repository here:  <https://github.com/kwartler/Vienna_July24> |

**Objectives and Learning Outcomes**

|  |
| --- |
| **By the end of the module students will have learned:**   * Introductory R syntax * Framing analytics and data science problems * Analytics workflow: Obtain, Scrub, Explore, Model, Interpret   **Knowledge and understanding**  After completing this module the students will have:   * Cleaned realistic data * Created visualizations and summary statistics to explore data * Fit a basic model to produce a patient’s probability for readmissions * Critical thinking skills to understand the implications of a model for not only organization value but also patient quality of care   **Cognitive and subject specific skills**  After completing this module the students will have the skills:   * To accomplish basic tasks with R * Basic computer science knowledge such as logic operations, loops, and functions   **Key skills**  After completing this module the students will:   * Confidently frame analytical projects * Have basic coding skills to accomplish some of the methods needed to complete analytical projects |

**Course Structure and Teaching Methods**

|  |
| --- |
| Each morning, the course will be delivered with a series of code examples and explanations where students can learn R syntax and appropriate steps of a data science project. Students will then divide into groups to accomplish the case with support from the instructor in the afternoon. At the conclusion of the two day seminar, groups will be randomly selected to present their code, and a business presentation. |

**Transversal Themes and Current Discourse**

|  |
| --- |
| **Contemporary issues highlighted in this module:**   * Responsible AI – the appropriateness of protected classes in/out of data IE gender, race etc. * Emerging AI regulatory landscape – EU AIA, NIST, AI healthcare regulations |

**Reading Material**

|  |
| --- |
| **Required readings**   * Chapter 1 of the free book: <https://intro2r.com/chap1.html> * Chapter 2 of the free book: <https://intro2r.com/basics_r.html> * Chapter 3 of the free book: <https://intro2r.com/data_r.html> * Chapter 5 of the free book: <https://intro2r.com/graphics_r.html> * Masterclass chapter on git & version control: https://intro2r.com/github\_r.html   **Additional readings**   * PDF of IBM Watson Oncology [here](https://github.com/kwartler/Harvard_DataMining_Business_Student/blob/master/EthicsArticles/IBM%20pitched%20Watson%20as%20a%20revolution%20in%20cancer%20care.pdf) regarding AI in healthcare. * FDA [article](https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device) on AI in medical devices * [Overview](https://haiweb.org/wp-content/uploads/2022/02/Interpreting-the-EU-Artificial-Intelligence-Act-for-the-Health-Sector.pdf) of the EU AIA for healthcare   **Recommended links and websites**   * Basic R syntax [cheatsheet](https://iqss.github.io/dss-workshops/R/Rintro/base-r-cheat-sheet.pdf) * Modeling “[cheatsheet](https://bcheggeseth.github.io/253_fall_2021/r-cheatsheet.html)” |

**Guidelines Regarding AI Tools**

|  |
| --- |
| AI tools, specifically large language models, ChatGPT and others, are encouraged and allowed for any assignment with proper citation. |

**Pre-Module**

|  |
| --- |
| **Assignment 1 (individual work):**  **Reading assignments/material**   * Install R, R-studio * Masterclass: install git, and lm-studio * Read Chapter1,2,3 from above * Review the FDA [article](https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device) on AI in medical devices * Review this [overview](https://haiweb.org/wp-content/uploads/2022/02/Interpreting-the-EU-Artificial-Intelligence-Act-for-the-Health-Sector.pdf) of the EU AIA for healthcare * [LM Studio Explanation](https://www.youtube.com/watch?v=yBI1nPep72Q&ab_channel=MatthewBerman) video * [R & Git Integration](https://www.youtube.com/watch?v=bUoN85QvC10&ab_channel=RiffomonasProject) Video   **Assignment 2 (individual work):**   * Write 1 paragraph summary of FDA article * Write 1 paragraph summary of EU AIA overview * Complete and submit the basic R code “[Intro\_to\_R\_Homework.R](https://raw.githubusercontent.com/kwartler/Vienna_July24/main/Intro_to_R_Homework.R)”   **Deadline for pre-module assignment 2 and process of delivery**  July 11, 2024 at 9:00am (CET) – upload on Moodle  *All student work is checked for both plagiarism and the use of AI upon submission.*  *Please submit your assignment as a PDF file onto Moodle, indicate your names in the document and name your document “SURNAME / GROUP NAME\_PRE”.* |

**Core-Module**

|  |
| --- |
| **Reading assignments/material**  Chapter 5 & 9 from above  Masterclass: video for lm studio explanation; git R studio integration  Obtaining data, cleaning, exploring, modeling and interpreting results of the healthcare data |

**Distinguished Guest Speakers**

|  |
| --- |
| Faculty instructions: In line with our commitment to promoting diverse perspectives and enriching the learning experience, we encourage you to invite guest speakers to your class - physically and/or virtually. If you do, kindly share their names and schedules for seamless integration into your teaching plan. Our team is committed to accommodating your needs.  Name:Tracy Ring, https://www.linkedin.com/in/tracyring/  Expertise and background: Chief Data Officer and Global Generative AI Lead - Life Sciences at Accenture  Topic: Responsible AI & Generative AI in business  Date and time: TBD (working on aligning her schedule and if she can be in person) |

**Detailed Schedule**

**Day 1**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | 08:30 – 10:30 | Environment set-up, R, R studio, Git (masterclass only), LM Studio (masterclass only)  Some R Coding examples | | 10:30 – 10:45 | *Coffee break* | | 10:45 – 12:15 | Framing an analytics project  Obtain data, clean/scrub data for modeling | | 12:15 – 13:15 | *Lunch time* | | 13:15 – 15:15 | Explore data – summary statistics, visualizations | | 15:15 – 15:30 | *Coffee break* | | 15:30 – 17:00 | Break into groups to start with case data performing Obtain, data cleaning and exploration steps to create a modeling matrix. | |

**Day 2**

|  |  |
| --- | --- |
| 08:30 – 10:30 | Two modeling methods, and evaluation |
| 10:30 – 10:45 | *Coffee break* |
| 10:45 – 12:15 | Improving model performance examples (ensemble, more observations, more features, feature engineering) |
| 12:15 – 13:15 | *Lunch time* |
| 13:15 – 15:15 | Group work to model, evaluate, interpret results then create slides for the case  Masterclass: Create a git repo for your group work  Masterclass: Using LM Studio to extract data from documents for use in modeling |
| 15:15 – 15:30 | *Coffee break* |
| 15:30 – 17:00 | Group presentations including code step through & slides explaining the modeling steps, results and implications |

**Post-Module**

|  |
| --- |
| **Assignment (individual work):**  New dataset, create a script that accomplishes all tasks Obtain Scrub Explore Model Interpret with a new dataset  **Deadline for post-module assignment and process of delivery**  August 5, 2024 @ 9:00am (CET) – upload on Moodle  *All student work is checked for both plagiarism and the use of AI upon submission.*  *Please submit your assignment as a PDF file onto Moodle, indicate your names in the document and name your document “SURNAME / GROUP NAME\_POST”.* |

**Assessment**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Pre-module:** |  | 30 % Intro to R Homework | | **Core-module:** |  | 35 % Group Work: Diabetes Data  Obtain Scrub Explore Model Interpret & slides | | **Post-module:** |  | 30 % Individual work: New data set script:  Obtain Scrub Explore Model Interpret [no slides] | | **Peer assessment:** |  | 5 % | |  |  | =100% | |

**Instructor and Contact Information**

|  |
| --- |
| **Ted Kwartler**  Organization: Accenture and Harvard Extension School  Position: Managing Director (North America Responsible AI Lead) and Adjunct Professor  E-mail: edwardkwartler@fas.harvard.edu |

**Short CV of Instructor and Contact Information**

|  |
| --- |
| Ted Kwartler is a distinguished data science and AI professional with roles in both academia and the corporate sector. He is an Adjunct Lecturer at Harvard University Extension School and Global School in Empirical Research Methods at St. Gallen University. After earning his Master’s of Business Administration from the University of Notre Dame, he served as an operations leader at Amazon.com, an AVP of Data Science Innovation at Liberty Mutual Insurance, followed by roles as the VP of Trusted AI and Chief Technology Officer for Generative AI at DataRobot. He's currently the Managing Director at Accenture charged with leading North American Responsible AI practice.  In addition to his corporate successes, Kwartler is recognized for his robust contributions to academia, developing innovative courses encompassing machine learning, natural language processing, and data-driven business decisions at renowned universities. His impact also extends to DataCamp.com, where his courses were taken by more than 32,000 in subjects including Text Mining, Sentiment Analysis, Statistics and Data-Driven Business Decisions. As a published author, Kwartler has penned important works on text mining, sports business analytics, and customer insight. In 2022, he was awarded Hult University's Faculty of the Year, and has been a member of the editorial board of the Applied Marketing Analytics Journal since 2017. |