**Program: Executive MBA**

**Specialization: Strategic Marketing & Sales**

**Syllabus for Module: Brand Management**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date:**  Nov 6-7 | **Lecturer:**  **Ted Kwartler** | **ECTS:**  2,5? | **Language:**  English |

**Course Description**

|  |
| --- |
| This module will take students beyond the basics of R programming and introduce them to more sophisticated techniques and tools for data analysis, modeling, and visualization. Students will learn how to apply their knowledge of analytics and data science to real-world business problems using R. They will be able to then take these findings and articulate them in a manner that executives and non-technical people can understand.  Course material will be on our class repository here:  <https://github.com/kwartler/Vienna_24> |

**Objectives and Learning Outcomes**

|  |
| --- |
| **By the end of the module students will have learned:**   * Advanced R syntax and programming concepts * Applying machine learning algorithms to solve complex business problems * Framing analytics and data science projects for business impact * How custom functions improve data scrubbing in R * Basic API parsing as a data source   **Knowledge and understanding**  After completing this module the students will have:   * The ability to apply decision trees, and random forest algorithms * Applied tuning parameters and knowledge of their impact to new modeling projects * Create customized R functions * Work with APIs using packages like `httr` and `jsonlite` in a modeling workflow * Use R to create interactive HTML pages for business users   **Cognitive and subject specific skills**  After completing this module the students will have the skills:   * To apply machine learning algorithms to complex business problems * Understand more advanced R programming and when to apply it * Think strategically about a business question and the appropriate data narrative, artifact or insight   **Key skills**  After completing this module the students will:   * Frame complex analytical projects for business impact * Customize R functions to meet specific business needs * Integrate R with other tools and technologies (e.g. APIs) * Apply more advanced algorithms on business data to improve insights and accuracy |

**Course Structure and Teaching Methods**

|  |
| --- |
| Lecture, group work, breakout sessions, hands-on labs |

**Transversal Themes and Current Discourse**

|  |
| --- |
| Faculty instructions: Please specify to students the contemporary issues that will be emphasized in your class, alongside the coverage of fundamental aspects in the field. Examples of such topics include, but are not limited to, sustainability, digital transformation, net zero, ESG, AI, diversity, equity & inclusion. This information will improve our course communication and ensure that students gain a comprehensive understanding of the course content.  **Contemporary issues highlighted in this module:**   * The ethics of algorithmic hiring, that AI fairness is complicated * Good AI Governance |

**Reading Material**

|  |
| --- |
| **Required readings**   1. **Data exploration and visualization in R**    * [Exploratory Data Analysis with R](https://r4ds.had.co.nz/exploratory-data-analysis.html) from R for Data Science by Hadley Wickham.    * [Data Visualization with ggplot2](http://r-statistics.co/Complete-Ggplot2-Tutorial-Part1-With-R-Code.html) from R Statistics. 2. **Regression modeling in R**    * The [R tutorial on Simple Linear Regression](https://www.r-tutor.com/elementary-statistics/simple-linear-regression): It provides a basic understanding of Simple Linear Regression and helps you to understand how to create a Simple Linear Regression model using R. The navigation is a bit odd for this but the red links at the bottom help you progress. 3. **Model evaluation in R**    * [How To Interpret R-squared and Goodness-of-Fit in Regression Analysis](https://statisticsbyjim.com/regression/interpret-r-squared-regression/) |

**Guidelines Regarding AI Tools**

|  |
| --- |
| Faculty instructions: Please specify whether the use of AI tools is allowed, partially allowed, or prohibited and outline the extent to which they can be used. For examples and guidance on how to articulate your approach to AI tools, please refer to the [EA Style of Teaching](https://teach.executiveacademy.at/en/) website. This information aims to ensure transparency and consistency in the use of AI tools throughout the course.  *Option 1*:  AI tools, such as ChatGPT and others, are allowed for any assignment with proper citation. |

**Pre-Module**

|  |
| --- |
| **Reading assignments/material**   1. Read the bolded best practices of visualization here: <https://www.tableau.com/blog/stephen-few-data-visualization> 2. Review <https://r-graphics.org/> to see the many types of charts available through ggplot2 3. Review the types of charts available at <https://echarts4r.john-coene.com/> 4. Review the syntax for creating interactive dashboard at <https://pkgs.rstudio.com/flexdashboard/>   **Pre-module activities**  All Tracks:  Choose the **basic** path if  1. have never programmed in R before.  2. you expect to manage or direct analytical groups with the goal of effective communication and collaboration  Choose the **advanced** path if  1. you have some R programming experience, in a professional setting or if applicable have a solid grasp of the previous session material  2. your goal is to become a Data Scientist yourself  Choose the **master** path if   1. you are proficient in R and have already used it for data analytics and/or machine learning purposes.   **Basic** Pre-Module Assessment:  Complete the following steps in R:   1. Install R & R-Studio   In the console (lower left by default) add packages used in class   * 1. `install.packages(“ggplot2”)`   2. Use the above function to install `ggthemes`, `dplyr`, `caret`  1. Watch the Introduction to R tutorial [here](https://www.youtube.com/watch?v=eR-XRSKsuR4). 2. Create a copy of the scaffolded script **`basic.R`** from our class repository. Fill it out and then save a copy to turn in.   **Advanced** Pre-Module Assessment:  Create a copy of the scaffolded script **`PreAssignment.Rmd`** from the class repository. Fill it out and then save a copy to turn in.   1. Complete the R script **`PreAssignment.Rmd`**. 2. Create 2 presentation slides    1. Slide 1: Make it a side-by-side slide:       1. Left side: Describe the machine learning workflow       2. Right side: Add a ggplot from the script or another you made while exploring the data    2. Slide 2: Make is a slide that has 2 sections side by side:       1. In each section: Name the model, create a 1 sentence “lay-person’s” description of the models’ differences, list the model’s KPI from the script.       2. Denote on the slide which of the two methods is the best model according to adjusted r-squared.   **Master** Pre-Module Assessment   1. Use the data **WA\_Fn-UseC\_-Telco-Customer-Churn.csv** to create a machine learning R script from scratch. You will be expected to follow these steps in your R script:    1. Sample – perform appropriate sampling of the data    2. Explore – Explore the training data building visualizations, summary statistics and identify any issues in the data.    3. Modify – Process the data to account for missing/NA, factors, and outliers.    4. Model – Perform **3** modeling techniques and tune appropriate parameters, and/or change included variables. The dataset is a “**churn**” prediction use case.    5. Assess – Assess each of the 3 modeling types to identify the best one. You need to choose an *appropriate* metric like RMSE, MAPE, LogLoss, Tweedie Deviance etc based on your understanding of the problem. 2. Create 2 presentation slides    1. Slide 1: Make it a side-by-side slide:       1. Left side: Describe the machine learning workflow with steps you’ve taken to clean the data *i.e. remove outliers if applicable*       2. Right side: Create a scatterplot of the Y variable with the most correlated x variable. Apply the g docs theme from `ggthemes`    2. Slide 2: Make is a slide that has 3 sections side by side:       1. In each section: Name the model, create a 1-2 sentence “lay-person’s” description of the model’s learning method, list any parameters or variables chosen to optimize the model output and list the model’s KPI from the script.       2. Denote on the slide which of the three methods is the best model.   **Deadline for pre-module assignment and process of delivery**  November 4, 2024 @ 9am (CET) via Moodle  *All student work is checked for both plagiarism and the use of AI upon submission.*  *Please submit your assignment as a n R file and pptx files (they can be saved a pdf too) onto Moodle, indicate your names in the document and name your document “SURNAME / GROUP NAME\_PRE”. The R Script should have a commented header with your name included and the slides should also have your name as a footer or elsewhere on the slide.* |
|  |

**Core-Module**

|  |
| --- |
| **Reading assignments/material**  Please review the AI ethics articles [here](https://github.com/kwartler/Vienna_24/tree/main/Fall_2024/ethics%20articles) as one shows an emerging aspect of AI ethics and another a longer standing use case that is fraught with potential bias.  In class all students will individually build a resume classification tool. They will need to then evaluate it for bias. |

**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: Peter Prettenhofer  Expertise and background: <https://www.linkedin.com/in/peterprettenhofer/>; contributor to sci-kit learn, previous VP Engineering at DataRobot, currently Principal Research Engineer Neo Cybernetica  Topic: Building an analytical career  Date and time: Day2 16:30-17:00 |

**Detailed Schedule**

**Day 1**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | 08:30 – 10:30 | Review, class repository set up (new students) git pull/clone  The SEMMA data science process  Decision Tree Explanation & Code  Random Forest Explanation & Code | | 10:30 – 10:45 | *Coffee break* | | 10:45 – 12:15 | Hands-on:  Challenge code along using “INNHotelsGroup.csv”  Challenge code along using “diamonds2023.csv” | | 12:15 – 13:15 | *Lunch time* | | 13:15 – 15:15 | Non-Traditional Investing Activity  Depending on time: defects\_data.csv  Introduction to Lending Club data | | 15:15 – 15:30 | *Coffee break* | | 15:30 – 17:00 | Lending Club group case work – Sample, Explore, Modify | |

**Day 2**

|  |  |
| --- | --- |
| 08:30 – 10:30 | Hands-on:Coding TBD AI Ethics Lecture  Hands on: Students will build a university admissions model & analyze it for bias as part of the core-assignment |
| 10:30 – 10:45 | *Coffee break* |
| 10:45 – 12:15 | APIs in analytical workflows  Echarts4r, r-markdown, flexdashboard to HTML pages |
| 12:15 – 13:15 | *Lunch time* |
| 13:15 – 15:15 | Lending Club group work continued – Model, Assess; time permitting make a flexdashboard of data insights |
| 15:15 – 15:30 | *Coffee break* |
| 15:30 – 17:00 | Hands-on:Challenge – use JinaAI to webscrape information from adidas, or another website.  Lending Club Group Work results  Guest Speaker: ~~Peter Prettenhofer~~ Alan Aqrawi |

**Post-Module**

|  |
| --- |
| **Reading assignments/material**  All tracks: Choose 3 articles from the <https://incidentdatabase.ai/>  **Post-module activities**  **All tracks:** Please submit a short (~2 paragraphs for each articles.) summary & reflection of each of the chosen incidents. Include basic mitigations or ideas to limit the problem cited in the articles.  Basic: Write an R script following a basic ML workflow to predict used car prices using this [file](https://raw.githubusercontent.com/kwartler/Vienna_24/refs/heads/main/Fall_2024/PostModuleAssignment/basic/newCars.csv) from our class repository. Make predictions on a test partition of 20% and identify the top 100 “best deals” meaning the most underpriced cars in the test partition data set. Turn in your error free R script covering all aspects of the SEMMA (or similar) ML workflow.  Advanced: Write an R script following an advanced ML workflow to predict used car prices from this [file](https://raw.githubusercontent.com/kwartler/Vienna_24/refs/heads/main/Fall_2024/PostModuleAssignment/advanced/newCar_w_outliers_more_NA.csv) which has more NA and outliers than the basic version. Make predictions on a test partition and identify the top 100 “best deals” of the partition. As an advanced practitioner you will perform significant EDA, remove any outliers, fix any NA’s, engineer new variables and train at least 3 models for assessment. You will then select the best model by KPI to identify the top deals.  Mastery: Same as the advanced project but now the files are still in their original JSON format. You will have to read in all JSON files and turn them into a single data frame for modeling. The folder of JSON files is here. In addition to the additional steps described in basic and advanced, create a dashboard of EDA visuals using R-markdown/flexdashboard and “knit” it to HTML. Additionally, you will need to write at least 2 custom functions and use a loop and/or lapply functions to perform the analytical workflow. Turn in your R script, markdown and HTML.  Hints and Tips:   * Make sure to submit the script with full output and no errors. * Use consistent code styling, organization of tasks and appropriate commenting * Make sure that your results are reproducible (i.e., it should be possible to run the notebook from top to bottom with identical results. * For Advanced & Master level – apply multiple modeling techniques, adjust variables and tuning parameters to improve results then compare model KPI to ensure the most robust model is providing predictions   Deliverables: All Tracks: A short 6 paragraph AI ethics paper covering 3 AI Incident Database articles  Basic: R script described above demonstrating all aspects of the workflow  Advanced: R script described above, CSV of top 100 best deals.  Master: R Script described above, markdown file and knit HTML dashboard  **Deadline for post-module assignment and process of delivery**  November 25, 2024 @ 9am (CET) via Moodle  *All student work is checked for both plagiarism and the use of AI upon submission.*  *Please submit your assignment as an R script, markdown, HTML and PDF slides as appropriate onto Moodle, indicate your names in the document and name your document “SURNAME / GROUP NAME\_POST”.* |

**Assessment**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | **Pre-module:** |  | 30 % (Basic, Advanced, Mastery Assignments) | | **Core-module:** |  | 30 % (Ethics Articles & Resume Classification) | | **Post-module:** |  | 40 % (“New Car” Assignment described above for your track) | |  |  | =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 the 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. |