

AAE Data Science Syllabus – DRAFT

Principle

- Aim to classically educate actuaries in data science up to the level where they can judge the merits or otherwise of professional data science work.

Commented [CF1]: Comment from Greg Doyle
The syllabus as presented is very comprehensive and represents an ambitious learning path for students undertaking the programme.

Overview – three main pillars

- Technical and practical skills
- Business and professional skills
- Ethical and human skills

Technical and practical skills

Three sub-pillars

- Data-ology – understanding modern data
- Data science methods and tools
- Deployment and visualization

Preliminaries

- Demonstrate an understanding of the following aspects of data science:
 - Concepts
 - Terms
 - Principles
 - Assumptions
 - Challenges
 - Limitations
 - Jargon and explain-ability

Data-ology

- Demonstrate an understanding of modern data, including:
 - Data storage technologies
 - Mobile technologies and cloud technologies
 - Training data, test and validation data sets
 - Data infrastructure
 - Data architecture
 - Data pipeline
 - Data scraping
 - Data wrangling
 - Data engineering
 - Databases, data warehouses, data lakes etc
 - Associated technologies
 - Other data concepts

Formatted: English (Ireland)

Formatted

Commented [CF2]: Suggestion from Greg Doyle

Formatted: English (Ireland)

Commented [CF3]: Comment from Greg Doyle
“You do not explicitly mention mobile technologies, cloud technologies (public, private, shared, cloud processing/data storage, e.g. IAAS, SAAS, PAAS, etc.), SQL, NoSQL and NewSQL (in terms of database technologies), which are important technologies for data scientists.”

Data Science Methods and Tools

- Data Science Methods and Tools
 - Demonstrate technical proficiency in the use of modern data science methods and tools
- Overview / learning objectives
 - What methods and tools exist?
 - Where they might be used?
 - When might they be used?
 - How do you use them?
 - Why might you use them?
 - What are the emerging tools and techniques?
 - What perspectives might be taken?
 - For example, modelling versus strategy
- Data Science Methods and Tools – Predictive Analytics / Machine Learning
 - Demonstrate technical proficiency in the use of Machine Learning / Predictive Analytics / Statistical Learning
- Machine learning (main topic)
 - Related topics, for example, data mining (discovery v prediction)
 - Potential benefits and limitations of machine learning
 - Concepts
 - Model training (estimation, assuming a target is defined)
 - Model validation (statistical testing, for example, goodness-of-fit)
 - Feature engineering and feature scaling
 - Regularization
 - Other
 - Approaches / ‘Types’
 - Unsupervised
 - Supervised (including semi-supervised)
 - Reinforcement / reinforced learning
 - Probabilistic / Bayesian approaches
 - Other types
- Machine learning algorithms
 - There is a long and growing list of different types of algorithms used in machine learning.
 - Examples...
 - Neural networks (different types basic, convolutional, recurrent, long short-term memory and deep learning)
 - Dimension reduction, including principal component analysis
 - Tree-based methods Support vector machines
 - Aspects of natural language processing Gradient boosted machines
 - Decision trees Random forests
 - Generalised linear models Support vector machines
 - K-nearest neighbour K-means clustering;
 - Regression (linear and polynomial regression) Classification trees
 - Clustering, including k-means Hierarchical clustering
 - Generative Adversarial Networks (GAN). Principal component analysis
 - Ensemble techniques GBM
 - Further algorithms to be added...

- Interpretability.
- Advantages and disadvantages and likely usage (and ease of use) of different algorithms.
- Emerging algorithms
- Data Science Methods and Tools – Software skills
 - Demonstrate technical proficiency in software skills and a wherewithal to make progressive and critical judgment regarding software for data science purposes.
 - Programming languages
 - R
 - SQL
 - Python
 - No code programming
 - Understanding necessary around code versioning (e.g. git and github), related software engineering concepts and code libraries.
 - Using and dealing with libraries created by other programmers/data scientists to create efficient, effective and successful implementations
 - Understanding the professionalism issues arising from reliance on the opinions of others (e.g. R packages).
- Data Science Tools – Modelling and Strategy
 - Demonstrate technical proficiency in the use of modelling and strategy in data science.
 - Modelling
 - Process modelling and model development and validation
 - For example, CRISP-DM
 - Strategy
 - What should we do, and why? (a potential competitive advantage for actuaries)
 - Necessary to:
 - Know limitations of any project
 - Know where you might need to hand over to an expert in one area
 - Know the context and the landscape and if it is uncharted territory
 - Know what the project will be used for and why
 - The narrative is key
- Data Science Tools – Emerging Methods and Tools
 - Demonstrate a wherewithal to keep up to date with emerging methods and tools in data science
 - Demonstrate a wherewithal to keep up to date with evolving methods, tools and other developments in data science.
 - Emerging methods and tools
 - For example, natural language processing (for example, ChatGPT)
- Other areas
 - Recommender algorithms Fraud detection
 - Image analysis and object detection Self-driving vehicles
 - Medical imaging and diagnostics Robotics and robo-advising
 - Hardware and the costs and economics arising

Formatted: English (Ireland)

Commented [CF4]: Comment from Greg Doyle: "I'm not sure what you mean by "Reliance on the opinions of others (e.g. R packages). ", however, using and dealing with libraries created by other programmers/data scientists is critical to efficient, effective and successful implementations."

Formatted

Formatted: English (Ireland)

Formatted: Indent: Left: 1.25", No bullets or numbering

Formatted: English (Ireland)

Commented [CF5]: From Greg Doyle "in terms of process modelling, you might mention some newer and emerging process models, however CRISP-DM is a seminal data mining/science process model that data scientists should be aware of, as many of the newer models are based on it."

Formatted

Commented [CF6]: Comment from Pedro: Rather than explaining why the strategy is what it is, actuaries should be able to contribute to / articulate a strategy designed to meet specified business goals. I hope this makes sense. Actuaries have a good understanding of the many levers in an insurance company, and their combined impact, including "secondary order effects" in Solvency, reserving, pricing, etc.

Deployment and Visualisation

- **Deployment and Visualisation**
 - Demonstrate a wherewithal to deploy the results of the application of data science methods and tools in a business environment
 - Demonstrate a wherewithal to ensure that the results of data science projects get operationalised/actually used in organisations.
- Demonstrate an understanding of data visualization techniques including emerging techniques.
- Actuarial applications
 - Regression for survival analysis in high dimension (penalized Cox model, Accelerated Failure Time models...)
 - Survival trees, survival random forests.
 - Neural networks for survival analysis
 - Other

Commented [CF7]: From Greg Doyle

1 Deployment of the results of data science projects is a common issue that data scientists need to improve on i.e. how to ensure that the results of data science projects get operationalised/actually used in organisations and you should consider including this in some way - possibly in the data science process modelling section. It is easy to develop models, but difficult to roll the outcomes/outputs into everyday use in companies.

Formatted: English (Ireland)

Business & Professional Skills – Data Science

- Regulation
 - Demonstrate understanding of the regulatory and legal frameworks governing the application of data science.
 - Demonstrate understanding of the political and cultural environment and trends that influence the application of data science.
 - Demonstrate an understanding of the trends and reasons for the trends driving changes in data science regulation.
 - Demonstrate an understanding of the environment impact of data science and likely regulation arising.
- Data science is increasingly regulated
 - For example, EU rules “AI for life” and “AI for good”.
 - Other examples, GDPR, DPC and DPOs
 - Demonstrate a capacity to compliance with national/EU regulations and an awareness of regulations (or lack thereof) in other jurisdictions.
- Business context
 - Demonstrate the application of general actuarial business context awareness to commercial use of data science.
 - Demonstrate an understanding of the new and emerging roles in data science
- Actuaries typically require a general business environment understanding, including
 - Stakeholders
 - Other aspects of the business environment covered elsewhere in the AAE Education Syllabus
- Roles in Data Science
 - New and emerging roles
 - Data Wrangler, Machine Learning Engineer, Data Ethicist & AI Ethicist and other specialist roles
 - Senior roles: Chief Data Officer, Chief Ethics Officer

Formatted: English (Ireland)

Commented [CF8]: This is an obvious addition and I'm a little embarrassed I didn't see the need for it before. I could be put under bullet point 2 (political and cultural environment) but it does need to be more explicit.

Commented [CF9]: Suggestion from Pedro: Comments from Pedro

You may want to consider the inclusion of an item regarding the environmental impact of data science. This relates to the energy consumption of training very complex models (how cleanly that energy is generated, how greater demand impacts the price individuals' pay for own consumption such as heating, etc). I personally think that it should be one of the factors considered when assessing what model(s) might be appropriate for a given task. Most complex models are currently trained using borrowed computing power (AWS, etc) so data scientists are not necessarily aware of the energy consumed, and its overall impact (<https://www.theguardian.com/technology/2023/jun/08/artificial-intelligence-industry-boom-environment-toll>) I think it falls within actuaries' duty to the wider public.

Formatted: English (Ireland)

Commented [CF10]:

From Greg Doyle
In the regulation section, you might consider mentioning GDPR, DPC and DPOs - <https://www.dataprotection.ie/> and compliance with Irish/EU regulations and have an awareness of regulations (or lack thereof) in other jurisdictions. I note that compliance is mentioned later in the syllabus.

- Client Relationships
 - Demonstrate an understanding of the principles behind building and maintaining client relationships.
 - Demonstrate awareness of examples of successful and unsuccessful client engagements.
 - Key element in the Alliance for Data Science Professionals Syllabus.
 - The world largely works based on human relationships rather than being based mainly on numbers.
- Communication and teamwork
 - Demonstrate a practical understanding of the communication skills required for successful business engagement in data science.
 - Demonstrate an understanding of the explain-ability challenges that arise in data science tools and methods, along with methods to overcome the challenges.
 - Demonstrate a practical capacity to work in a team, including an understanding of the challenges that occur and methods to overcome them.
 - Communication skills with colleagues and with the general public are both imperative.
- Innovation skills
 - Demonstrate the creative use of actuarial skills.
 - Demonstrate a capacity to take concepts from outside of the financial/insurance industry and use them in the financial/insurance industry and vice versa.
 - Demonstrate a capacity to evolve existing methods and tools in data science.
- Project management
 - Demonstrate an understanding of the principles involved in project management in data science.
- Business leadership / C-suite skills
 - Demonstrate an understanding of the skills required for senior positions of responsibility in data science in business environments, for example: chief data officer and chief ethics officer.
- Application to actuarial problems
 - Demonstrate a capacity to apply modern data science methods and tools to actuarial problems.
 - Examples:
 - Regression for survival analysis in high dimension (penalized Cox model, Accelerated Failure Time models...)
 - Survival trees, survival random forests.
 - Neural networks for survival analysis
 - Other

Ethical & Human Skills – Data Science

Narratives - Turning Data into a Story

- Narratives
 - Demonstrate a wherewithal to make progressive judgment on ethical narratives related to data science.
- Turning data into a human narrative
 - Narrative discovery / assessment / development
 - Also know as “storytelling”
- Narratives dominate and limit analyses
 - Need to be able to engage at the narrative level to act in the public interest.
- Classical narratives (things as they are) versus romantic narratives (as someone might want things to be).

Character - Understanding the People Involved

- Human side to data science
 - Demonstrate a wherewithal to make progressive judgment on human matters related to data science.
- Character:
 - Assessing, managing and developing the people involved in the narrative
 - Character assessment
 - Character development
 - Assessing data science from a human perspective.

Who Can You Trust?

- Trust
 - Demonstrate a wherewithal to assess and to achieve trustworthiness
- Who, what and by how much can be trusted?
 - Assessing trustworthiness
 - Achieving trustworthiness
- Trust and professionalism
- Key element in achieving efficacy
 - Key element in the Alliance for Data Science Professionals Syllabus.

Ethical Principles

- Demonstrating an understanding of the following ethical principles and constructs.
 - Tradition v Progress
 - Progress
 - Principles of progress
 - Progressive and regressive constraints on progress
 - Cultural
 - Behavioral
 - Prudence
 - Responsibility (individual & social)

- Demonstrate a wherewithal to make judgments based on these ethical principles and constructs.

Compliance

- Compliance
 - Demonstrate a wherewithal to make progressive judgments regarding adhering to and respecting current cultural traditions and regulations in data science.
- Imperative of understanding existing norms and achieving compliance with professional standards.
 - Examples for data ethics
 - Discrimination / fairness
 - Bias, gender or ethnic
 - Privacy

Excellence

- Excellence
 - Demonstrate a wherewithal to transgress norms to create more progressive norms within the application of data science

Ethical leadership

- Ethical leadership
 - Demonstrate a wherewithal to transgress norms to create new, more progressive and socially responsible norms in the application of data science

<END>