Institute
and Faculty
of Actuaries

Communicating Data Science

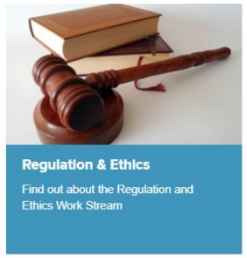
Alex Labram
Stelio Passaris

15 September 2020

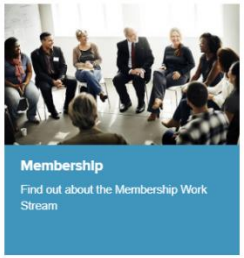
Data Science Working Party – Managing Committee

The overall objective of the Data Science Working Party is to be a platform of delivering Case Studies, Webinars, Events, GIRO sessions and integrating data science applications within our IFoA educational system, in order to supply actuaries and data practitioners with credible techniques that can be used within industry.

- Asif John (Chair)

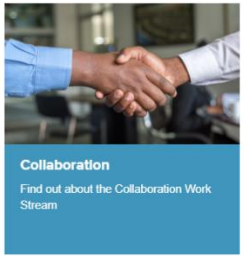


Matthew Byrne



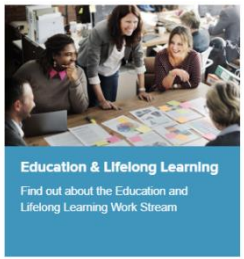
Alex Labram

Stelio Passaris



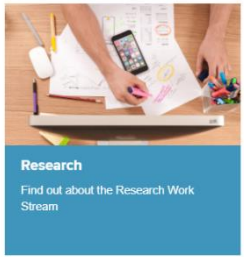
Biren Dhir

Alexis Iglauer



David Tudor-Griffith

Ankush Agarwal



John Ng

New COVID-19

Managing Committee

Institute and Faculty of Actuaries

Data Science Working Party – Membership Workstream

The Membership workstream supports a dialogue between the Data Science Working Party and the wider membership of the IFoA, both to inform members and to promote data science within the Profession.

Alex Labram

Alex is a manager in professional services firm Grant Thornton's Actuarial & Risk team, where he specialises in quantitative modelling, information systems and deriving insight from data.

Alex has over a decade of experience across the world of data analytics, particularly the financial sector. His diverse consulting background has given him exposure to a wide range of client types, problem domains and solution architectures.

Stelio Passaris

Stelio Passaris is a systems actuary and solution architect, specialising in actuarial modelling, process improvement, and system automation.

He is a Senior Actuarial Manager at FIS, the world's largest financial software company, where he provides consulting support on Prophet, the global leader in actuarial projection software.

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Background

Agenda

- Introduction
- Professional, legal and ethical requirements
- Elements of data science
- Components of a communication strategy
  - Communicating results
  - Communicating methods
  - Communicating implementations
- Wrap-up

Institute and Faculty of Actuaries

Professional, legal and ethical requirements

- Due to its novelty and power, Data Science usage must be carefully controlled and constrained
- Professional conduct extended for Data Science adaptation
  - IFoA: Actuaries' Code (section 6)
  - FRC: Technical Actuarial Standard 100 (sections 2.3-2.5, 3.2-3.5, 4.1-4.5, 5)
- "A Guide for Ethical Data Science" by RSS and IFoA
  - Seek to enhance the value of data science for society
  - Avoid harm
  - Apply and maintain professional competence
  - Seek to preserve or increase trustworthiness
  - Maintain accountability and oversight

https://www.actuaries.org.uk/upholding-standards/data-science-ethics

Institute and Faculty of Actuaries

Professional, legal and ethical requirements

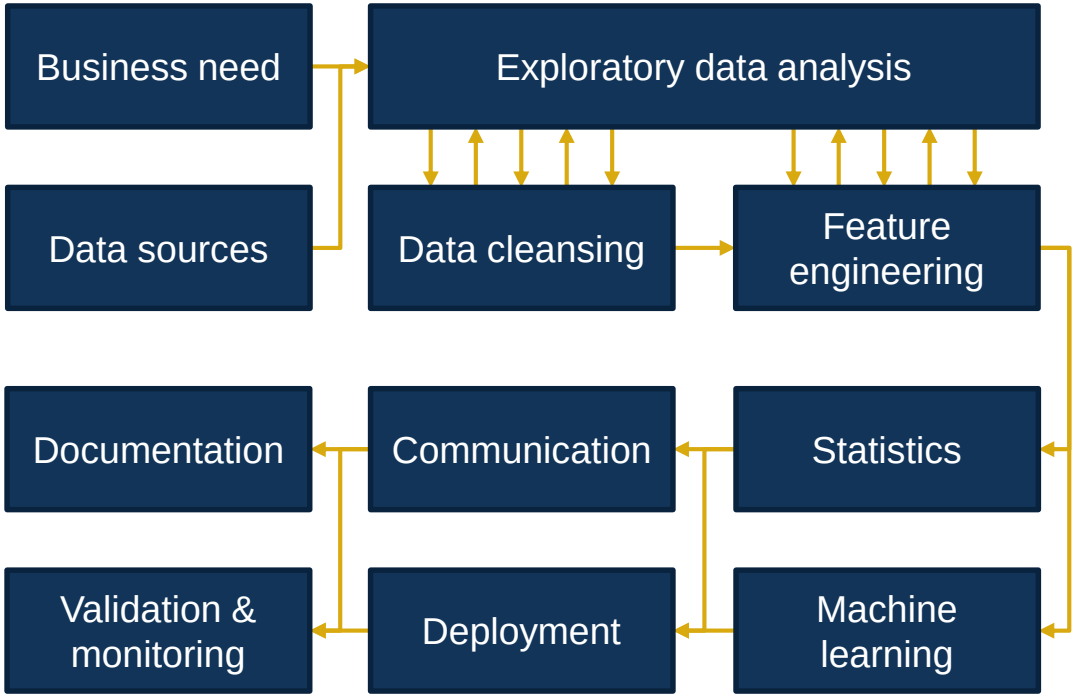
- Legal
  - Third-party sources used for analysis should be legally (not only physically) available
    - no unauthorised web-scraping
  - Protected characteristics must be treated with care
    - gender-specific pricing is banned in EU since 2012
  - Legal responsibilities and fines for personal data leakages
    - see GDPR
  - Regulatory obligations should be satisfied
    - full transparency on the algorithms applied
  - Insureds' profiling can bring additional issues
    - compatibility with GDPR requirements
    - information asymmetry reduction leading to diluted insurance risk definition

https://www.actuaries.org.uk/learn-and-develop/online-learning-resources-video-and-audio/professional-skills-training/trusted-profession

Institute and Faculty of Actuaries

Elements of data science

The data science process



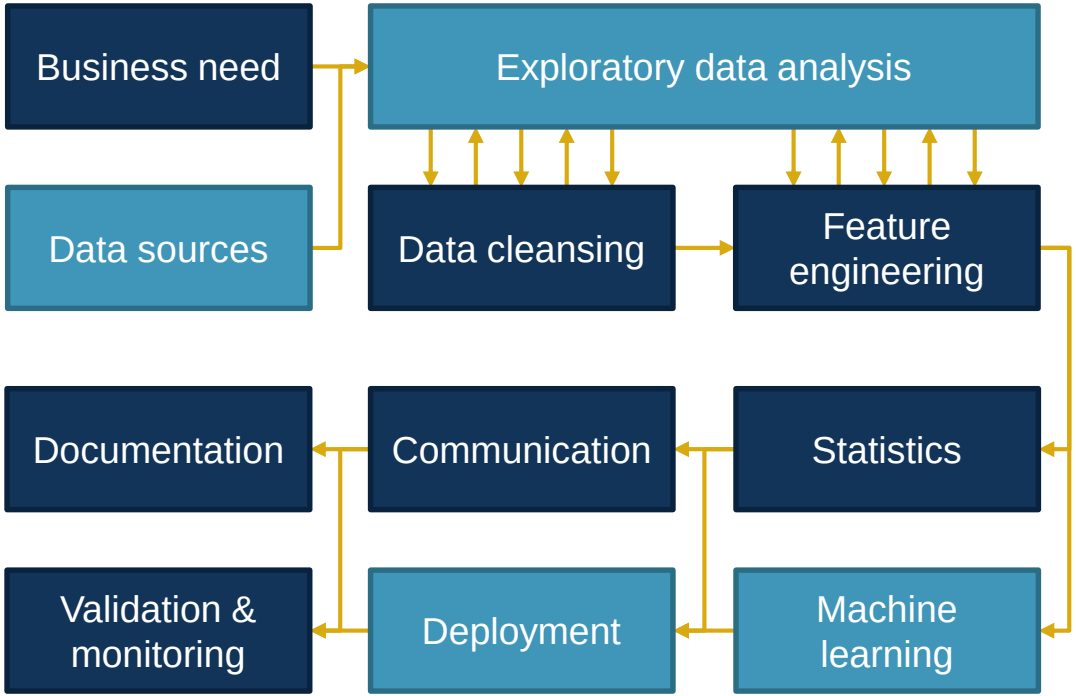
https://www.actuaries.org.uk/news-and-insights/news/data-science-process

"[A] multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data"

Institute and Faculty of Actuaries

Elements of data science

Differences to normal actuarial practice



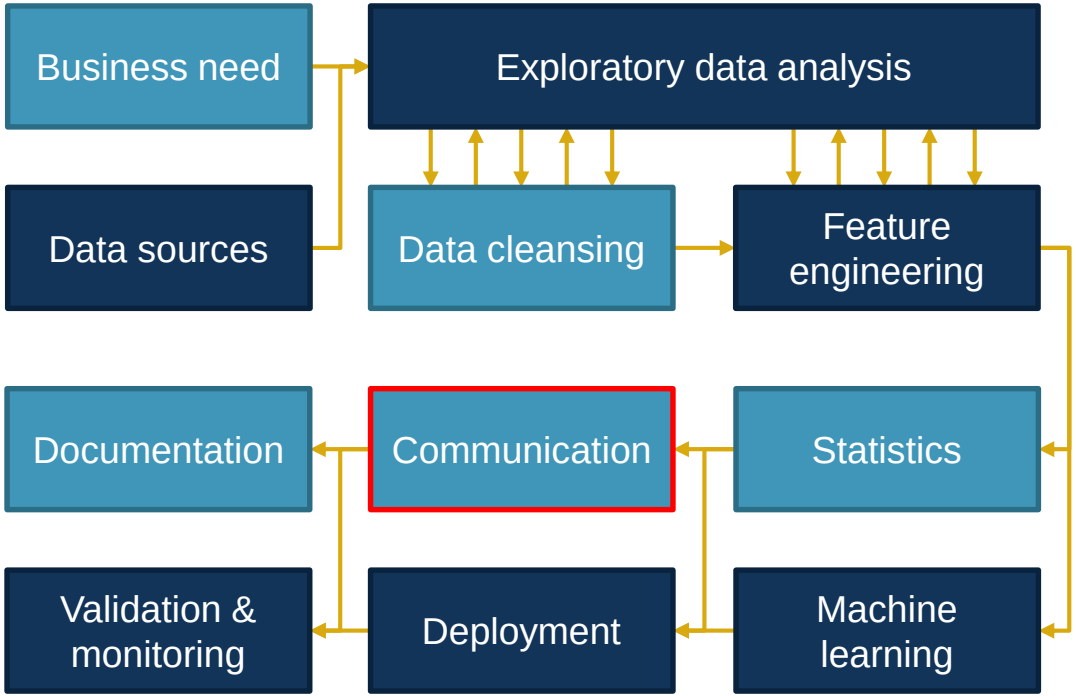
- Datasets
- Toolsets
- Skillsets
- Mindsets

https://www.actuaries.org.uk/news-and-insights/news/what-makes-data-scientist

Institute and Faculty of Actuaries

Elements of data science

Similarities to normal actuarial practice

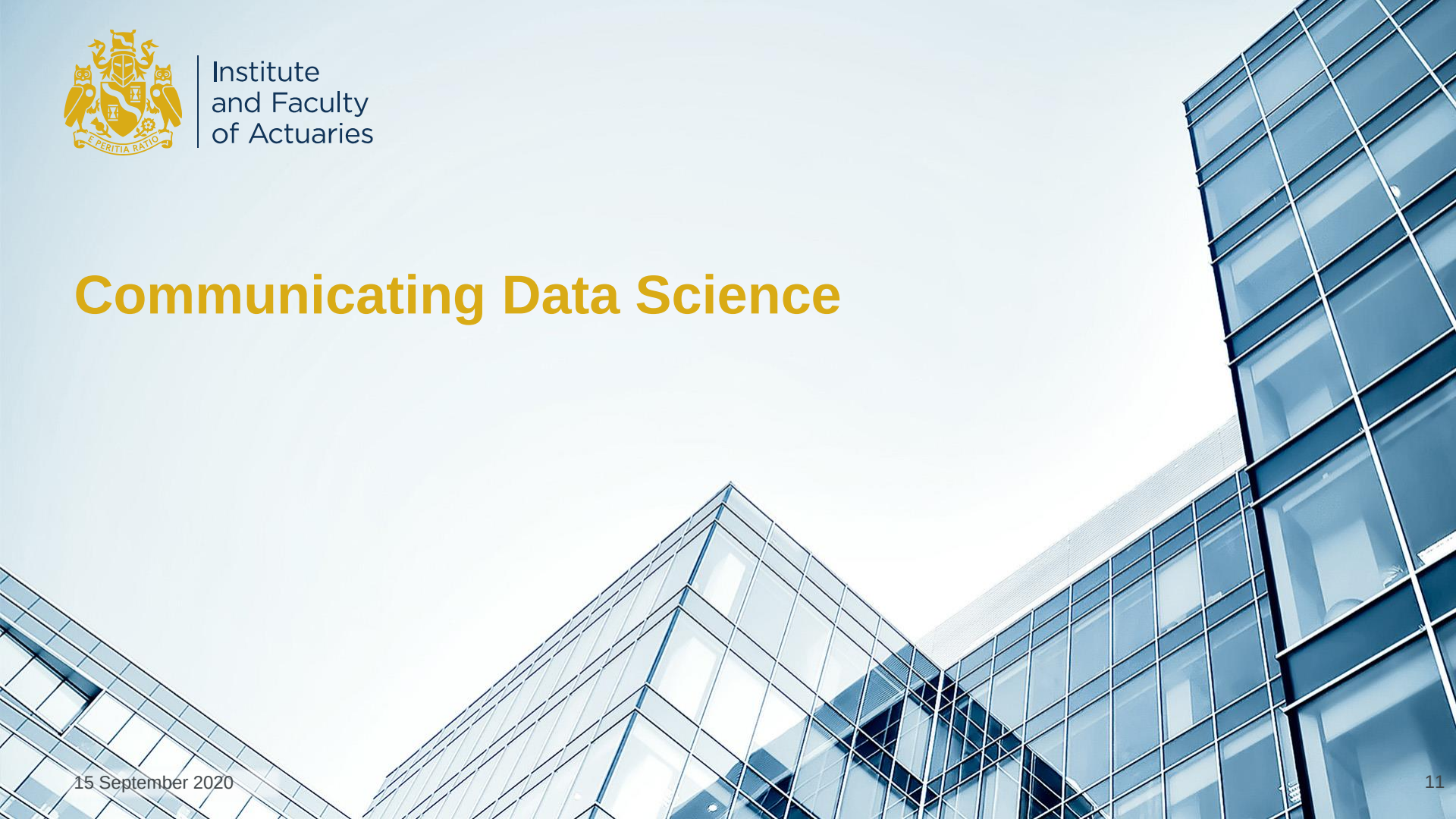


- Need for business/industry awareness
- Statistical foundations
- Emphasis on communication…
  - …but a broader range of options available for this!

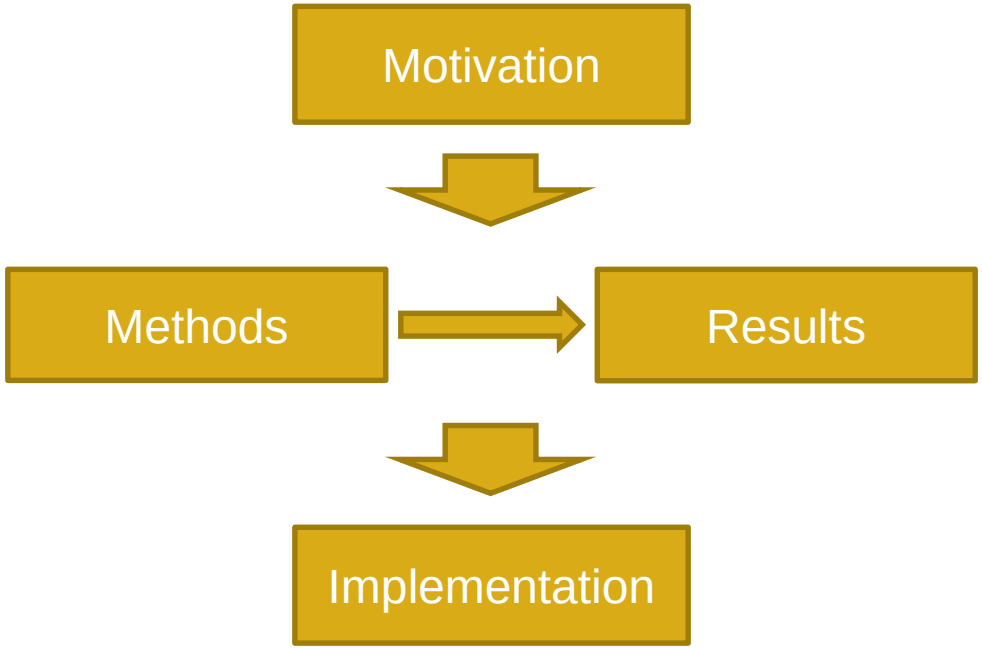
Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Communicating Data Science



Components of a communications strategy



- What – topic, granularity
- Why – purpose
- Who – audience
- How – medium

Institute and Faculty of Actuaries

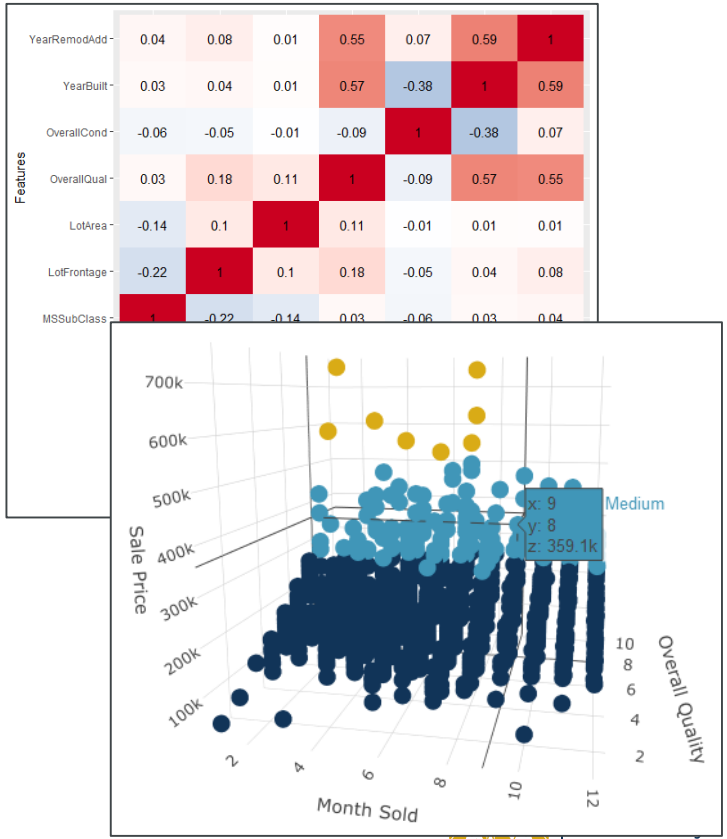
Communicating results
Key considerations

- Audience: business users
- Goal: communicate outcomes and limitations of analysis
  - “accurate, not misleading, and contains an appropriate level of information” (Actuaries’ Code 6.3)
- What your data scientist may forget
  - Interim and incidental findings
  - Uncertainties and sensitivities
  - Comparison with previous results
  - Reasonableness and smoothing
  - Ethical considerations
- Communication tools
  - Data visualisation (“viz”)
  - Model introspection / explanation
  - Online models

Institute and Faculty of Actuaries

Communicating results
Data viz

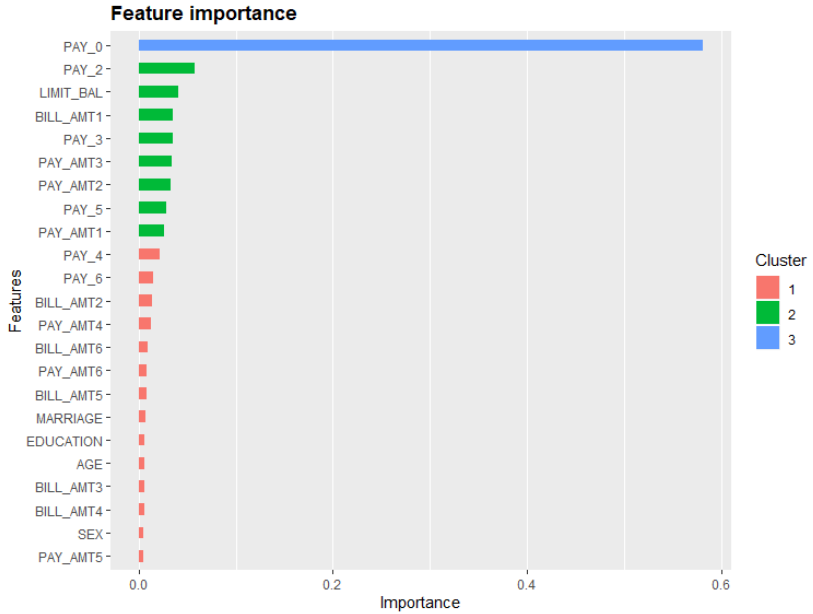
- One of the fastest ways to communicate strong messages
- Works well through all the project stages :
  - Initial data exploration and idea generation
  - Prominent showcasing of draft ML model output
  - Wrapping technical deliverables for digestion by decision-makers
  - Dashboards enable non-technical users to interact with data
- Depending on the user's skills, a few options are available:
  - Zero coding with stand-alone platforms (Power BI, Tableau)
  - Various programming packages (Seaborn, Dash, Bokeh, Plotly)



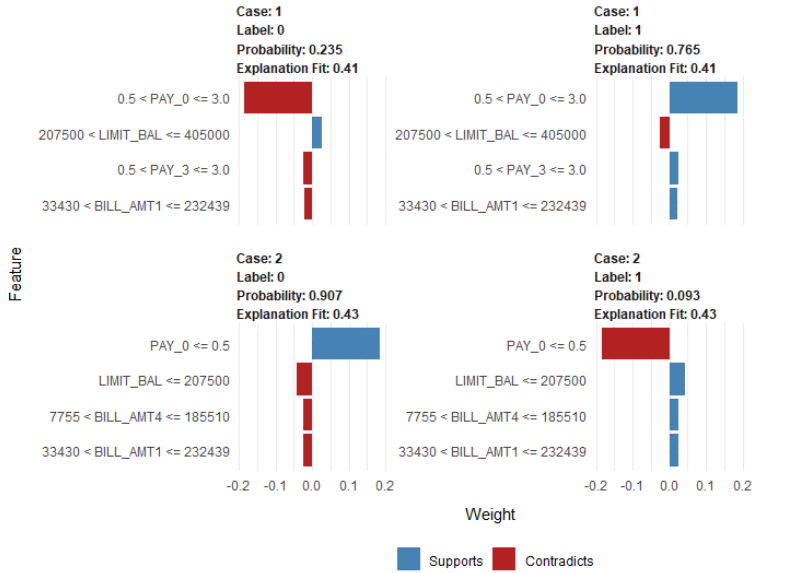
https://www.r-graph-gallery.com/

Communicating results
Model introspection

Model-specific methods



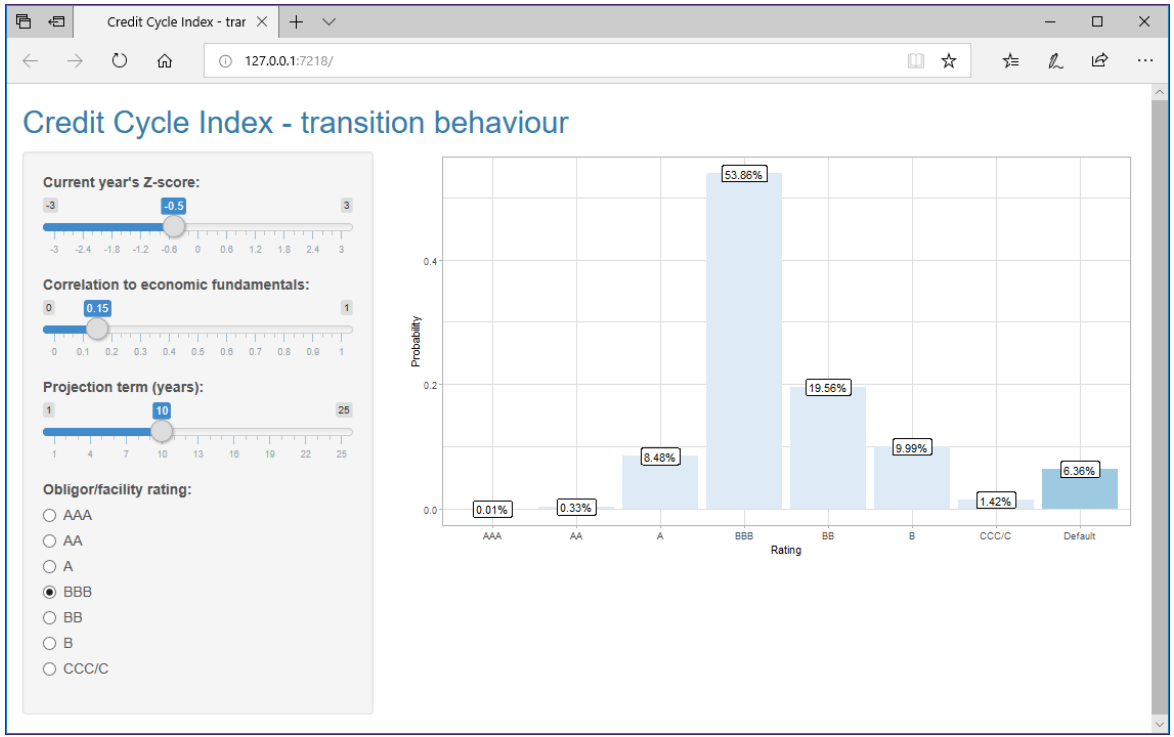
Model-agnostic methods (SHAP, LIME)



https://christophm.github.io/interpretable-ml-book/

Communicating results

Online models



Institute and Faculty of Actuaries

Communicating methods
Key considerations

- Audience: peers – actuaries, data scientists
- Goal: replicability

- Common issues
  - Different emphasis: assumptions vs approach
  - Post-hoc justification of ad-hoc exploration
  - Difficulty of explaining ML methods

- Communication tools
  - Model cards

Institute and Faculty of Actuaries

Communicating methods
Documenting ML methods

	Choice	Strengths & weaknesses
Algorithm family	• Tree ensemble	• Performs well across a range of problems • Non-parametric - handles biased / skewed data well • Handles interaction terms well • Handles proportional responses poorly
Algorithm	• XGBoost	• High-performing on small / medium datasets • Handles missing data well • Prone to overfitting if used without good validation
Target feature	• Burning cost as a fraction of maximum sum insured	• Burning cost is key variable for policy pricing • Taken as % to SI to eliminate proportionate response
Validation strategy	• 70:30 training / hold-out split • Fivefold cross-validation	• Data split broadly appropriate for small / medium dataset • No stratification therefore more vulnerable to idiosyncrasies in hold-out dataset
Accuracy metric	• Mean absolute error	• Appropriate for continuous outcomes; inappropriate for categorical outcomes • Robust to extreme outliers
Tuned hyper-parameters	• Learning rate (eta): 0-1 • L2 regularisation (lambda): 0-200 • Tree size (max_depth): 1-20	• Reduced risk of over-shooting / under-shooting best model • Better handling of collinear data • Limit over-fitting of overly complex apparent patterns

https://arxiv.org/abs/1810.03993 - Model Cards for Model Reporting
https://www.actuaries.org.uk/news-and-insights/news/article-fitting-data-xgboost

and Faculty of Actuaries

Communicating implementations

Key considerations

- Audience: maintainers / (re)users
- Goal: reproducibility

- What your data scientist may forget
  - Environment properties
  - Language and package versions
  - Random seeds
  - Deployment mechanisms
  - Post-deployment monitoring

- Communication tools
  - Literate programming – e.g. R Notebooks
  - Version control systems

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Wrap-Up

Questions

Comments

Institute and Faculty of Actuaries