



An Actuarial Engineering Perspective: Why Actuarial Science and IT Skills must Converge

Xavier Maréchal, Reacfin

About the speaker

- **Xavier Maréchal**– *CEO Reacfin and IA/BE qualified actuary, Reacfin*

Expert in Non-Life and Health insurance (pricing, product development, reserving and risk management) and Data Science.



Reacfin

We offer consulting services in actuarial science & quantitative finance, including a.o. capital - portfolio - product - risk - and liquidity - management. We build our expertise on broad data science capacities.

By blending strong actuarial and financial business expertise with an in-depth understanding of cutting-edge IT technologies, Reaxii enables our clients to become more competitive and focus on their core business such as complex analysis, strategic decision-making and innovation.

We share our knowledge with our clients. We offer a comprehensive learning platform, including on-site trainings, e-learning modules, webinars etc.



AGENDA



- (IT) Skills evolution for actuaries
- Creative data sourcing and management
- Advanced predictive modelling
- Upgraded coding, operationalization and reporting skills
- Conclusion

Skills evolution for actuaries

What are the new IT skills actuaries should develop?



Data Engineering (ETL from various sources, understanding data pipelines)	Machine learning and AI (AI, ML)	Automation and Scripting (Python, PowerShell, automation tools)
DevOps Practices (Collaboration between development and operations teams)	Blockchain technology (Decentralized and secure transactions)	Cybersecurity awareness (Risks and best practices)
Database management (Database management systems, SQL)	Data security (Encryption, access control and secure data storage)	Big Data Technologies (Apache Hadoop, Spark)
IT Governance and Compliance (Familiarity with frameworks)	Cloud computing (AWS, Azure, Google Cloud)	API integration (Data exchange)



Skills evolution for actuaries



What are the IT-oriented skills actuaries should develop? → My view 😊

Data Engineering (ETL from various sources, understanding data pipelines)	Machine learning and AI (AI, ML)	Automation and Scripting (Python, PowerShell, automation tools)
DevOps Practices (Collaboration between development and operations teams)	Data Visualisation (Tableau, PowerBI)	Cybersecurity awareness (Risks and best practices)
Database management (Database management systems, SQL)	Advanced Excel skills (Improve efficiency in data manipulation and analysis)	Ethical considerations (Data use, privacy and transparency)
IT Governance and Compliance (Familiarity with frameworks)	Cloud computing (AWS, Azure, Google Cloud)	API integration (Data exchange)



AGENDA



- (IT) Skills evolution for actuaries
- **Creative data sourcing and management**
- Advanced predictive modelling
- Upgraded coding, operationalization and reporting skills
- Conclusion

Creative data sourcing and management

Avoid the ""shit in, shit out"" effect



Data Engineering
(ETL from various sources,
understanding data pipelines)

Data Visualisation
(Tableau, PowerBI)

Database management
(Database management systems,
SQL)

Ethical considerations
(Data use, privacy and transparency)

API integration
(Data exchange)

Creative data sourcing and management

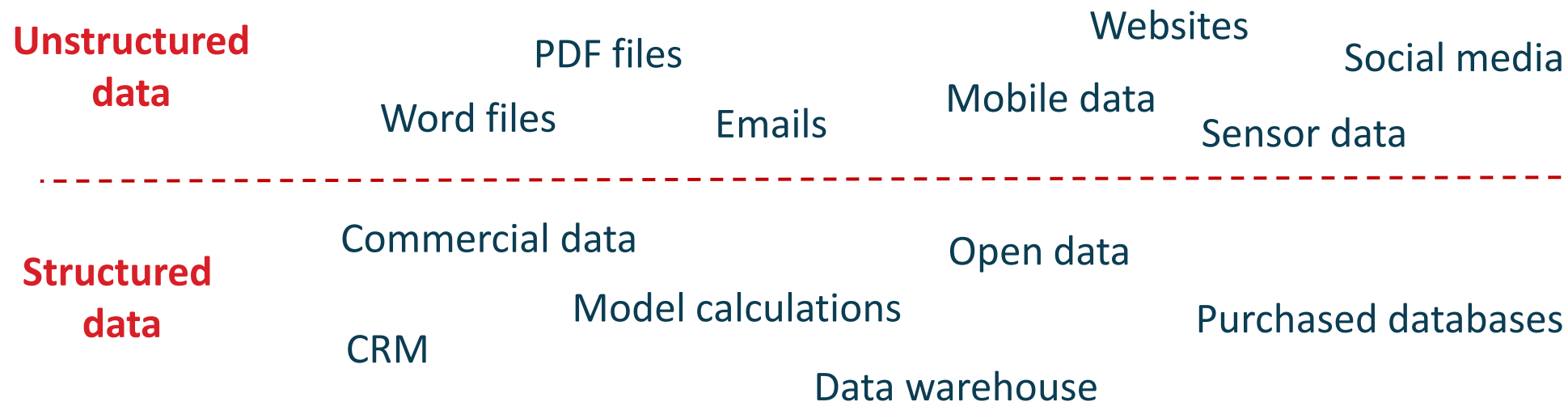


There exist 2 main categories of data

- **Structured data :**
organized and well characterized data that are easy to use because they are well identified.
 - E.g. insurer's policies and claims data
- **Unstructured data:**
non-organized data not easy to manipulate and which require much preparation (everything else).

80%

Of business information are unstructured



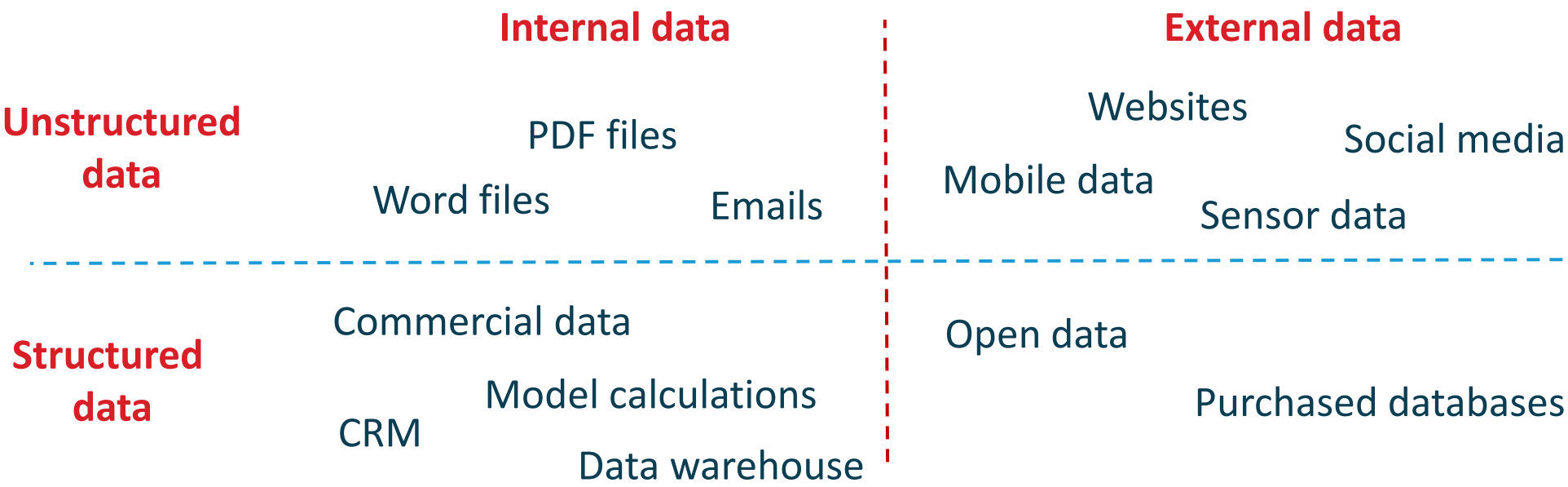
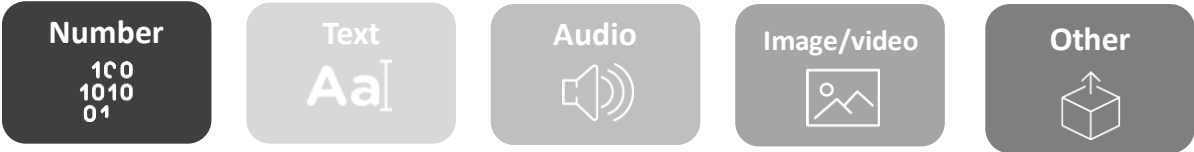
Creative data sourcing and management



Different sources and types of information

- Numerous sources of internal or external data
- Data type is different from one content to another

Increasing complexity
to collect and manage
unstructured data



Creative data sourcing and management

Data Engineering
(ETL from various sources,
understanding data pipelines)

Database management
(Database management
systems, SQL)

API integration
(Data exchange)



Data sourcing

- Additional data can be obtained through **many different sources** :

- 1. Scraping/parsing techniques:**

- Extract information automatically from websites

- 2. Open data files:**

- Structured datasets available to everyone

- 3. IoT sensor and API technologies:**

- Connected objects and application programming interface

- 4. External data providers**

- 5. Look twice into your own unstructured data:**

- Reveal hidden information from core data



Thanks to their business knowledge, actuaries should be leading the identification and collection of additional relevant data

Creative data sourcing and management

Data Visualisation
(Tableau, PowerBI)

Data Engineering
(ETL from various sources,
understanding data pipelines)



Data engineering

- Once additional data has been collected, **new methods and algorithm** allow to get the most out of it. Among others:

1. Statistics, ML and feature engineering:

Create structured features using initial data sources or charts to understand data

2. Text mining, NLP and LLM

Process of examining large collection of written resources and methods to perform linguistic analysis and/or generation

3. Image processing

Techniques to perform operations on images to enhance its content or extract information

Importance for actuaries

Actuaries should understand and make the most of these data engineering techniques to complete their traditional databases

Creative data sourcing and management



External and new data can be used to enrich the existing database with new attributes/variables

- 2 different points of view

- Data Scientist and Actuaries: enrich the existing database with a set of features which will be used when calibrating the models.

More variables should lead to better predictive models (! Let's avoid overfitting !)



- Business users (e.g. underwriter and marketing teams): **simplify** the models' results interpretation and/or the processes

- e.g. reduce forms sizes in underwriting → Quick quote

- What happens if too much data?

- Need for **feature selection**: keeping only the most relevant variables
- Potential for **feature engineering**: creating new variables to solve our problem

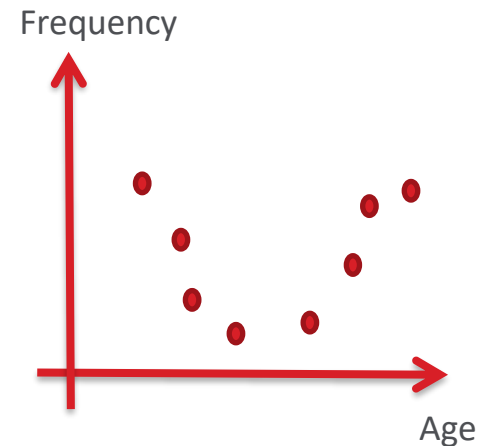
Creative data sourcing and management



Feature Engineering

- Features Engineering is absolutely known and agreed to be key to success in applied machine learning.
- Features Engineering is a Representation Problem
 - Machine learning algorithms learn a solution to a problem from sample data.
 - In this context, feature engineering asks: what is the best representation of the sample data to learn a solution to your problem?

Age	Frequency
-	-
-	-
-	-
-	-

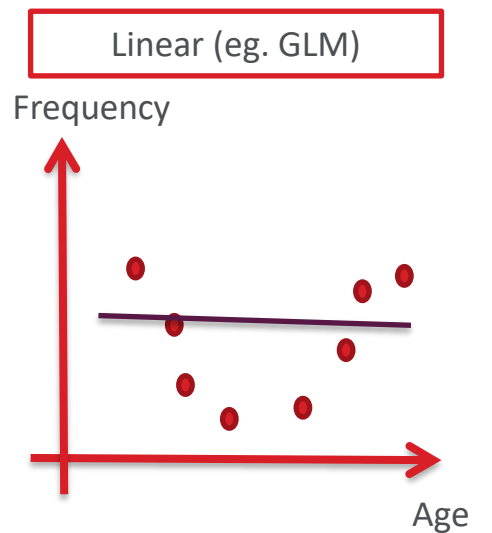


Creative data sourcing and management

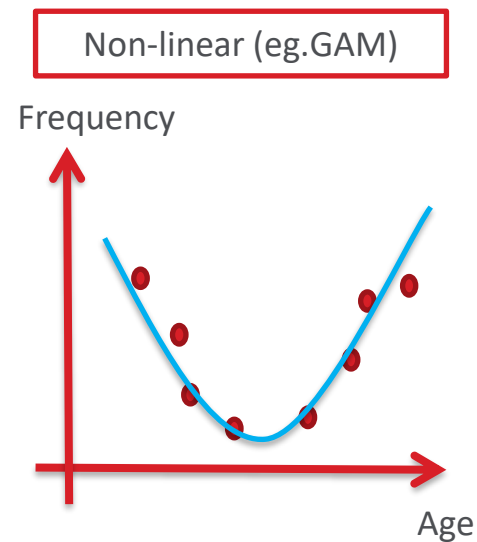


Feature Engineering vs Model Complexity

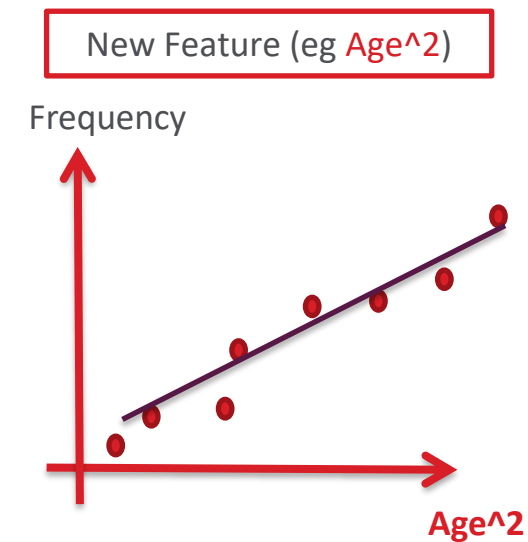
- The results you achieve are a factor of :
 - the model you choose,
 - the data you have available
 - and the features you prepared.
- The better the features that you prepare and choose, the better the results you will achieve



Poor Model choice and without features engineering



More complex model and without features engineering



Simpler Model choice BUT features engineering

AGENDA



- (IT) Skills evolution for actuaries
- Creative data sourcing and management
- **Advanced predictive modelling**
- Upgraded coding, operationalization and reporting skills
- Conclusion

Advanced predictive modelling

Machine learning offers new opportunities for predictive modelling but must be handled with care to avoid overfitting and/or black-box effect



Machine learning and AI
(AI, ML)

Data Visualisation
(Tableau, PowerBI)

- Results of Machine Learning algorithms will need careful attentions as they derive from automated procedures and **could induce conclusions which do not match a business logic**.
 - **Interpretability** is essential to avoid the black-box effect
- Another key challenge with Machine Learning is the risk of **overfitting**.

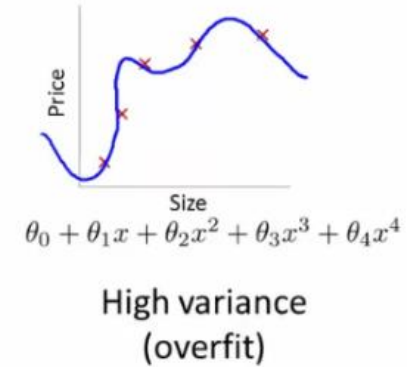
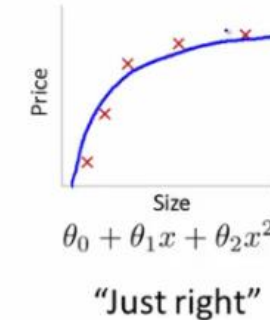
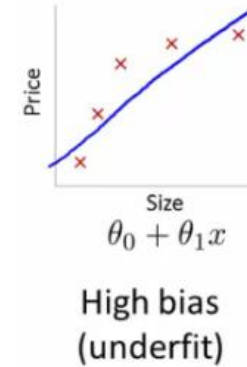
Overfitting

Overfitting deteriorates the predictive power of the model

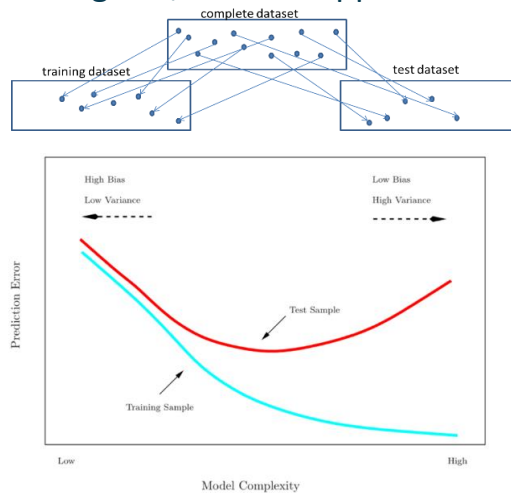
The overfitting problem

- When modelling, we should be sensibilized with overfitting/lack of parcimony.
- It occurs when a statistical model describes random error or noise instead of the underlying relationship.
- The fact that the model fits our data well doesn't guarantee it will be a good fit to new data
→ A good model is one that fits also well new data, i.e. that has a small predictive error

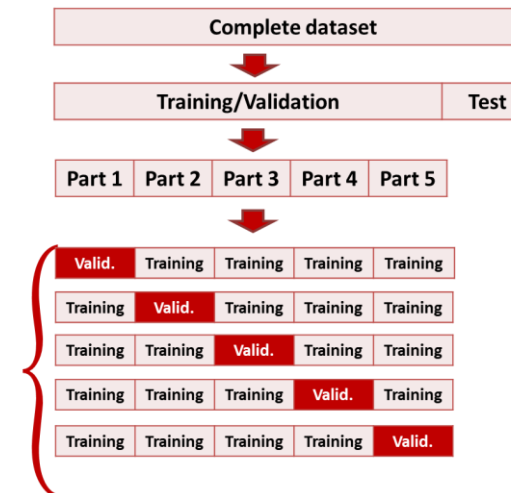
Potential solutions



Training set / Test set approach



Cross-validation

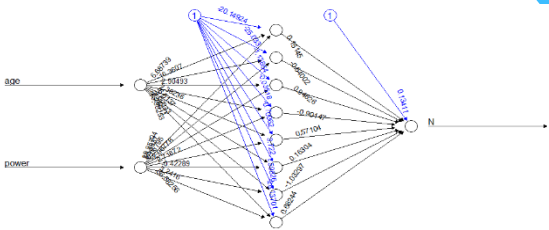


Some machine learning techniques are black boxes and interpretation of the results can be quite difficult

Understanding the results of ML techniques is not easy. Therefore, using explainable ML techniques is a must!



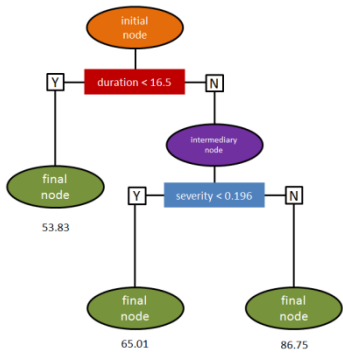
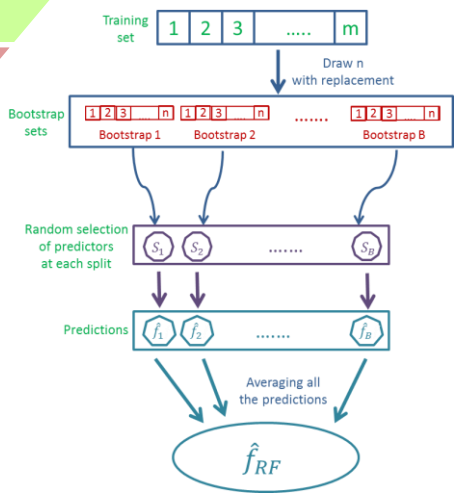
Complexity



In the case of regression trees, understanding how the model predicts claims' cost or frequency values for new data points is not a problem, as it is very intuitive.

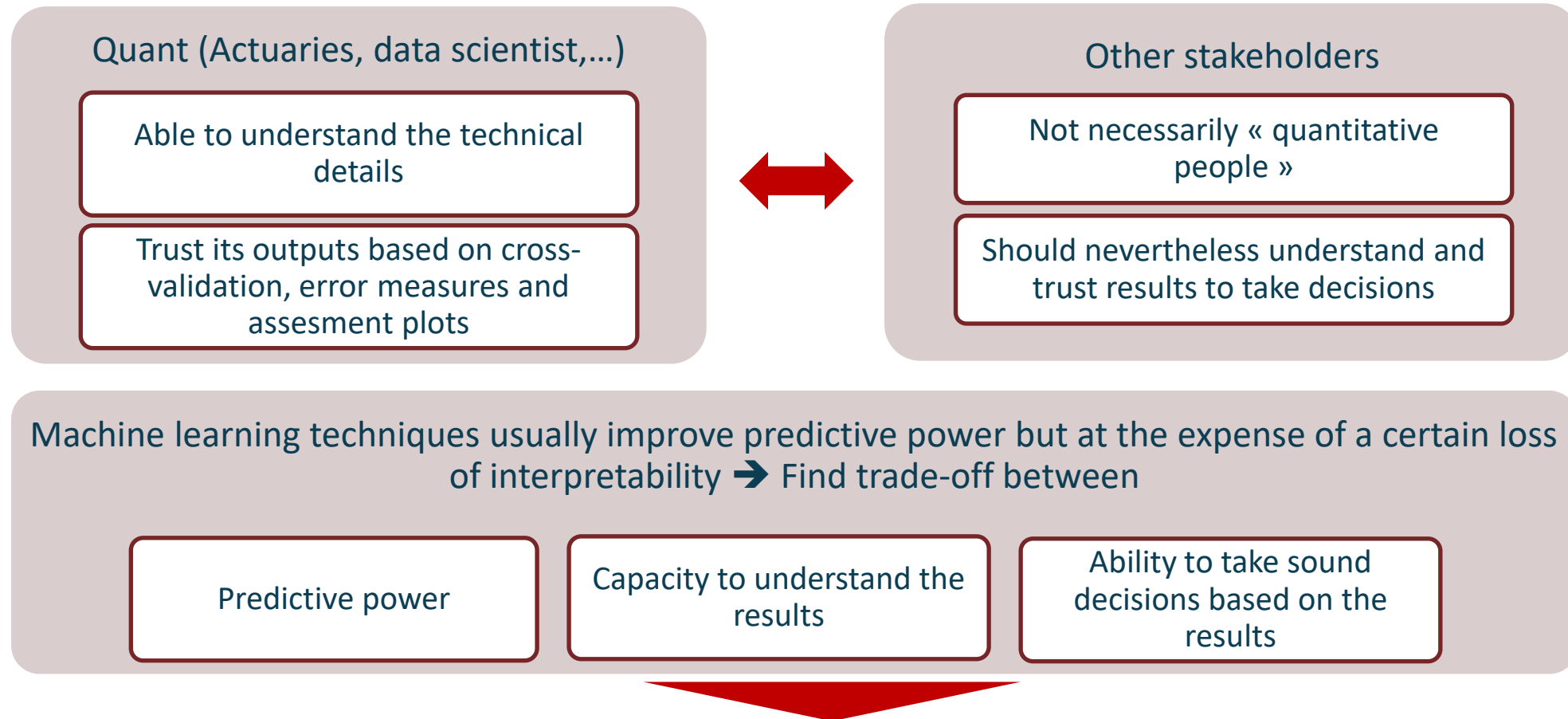
In the case of more complex methods such as Bagging and Random forests, even understanding how the model predicts values for new data points is rather difficult.

Things may be even worse for GBM and ANN.



Interpretability

Understanding the results of ML models is nevertheless key for sound business decision-making as many stakeholders use the results of the models



High-end questions

Who will use the results? For what purpose? With which impact?

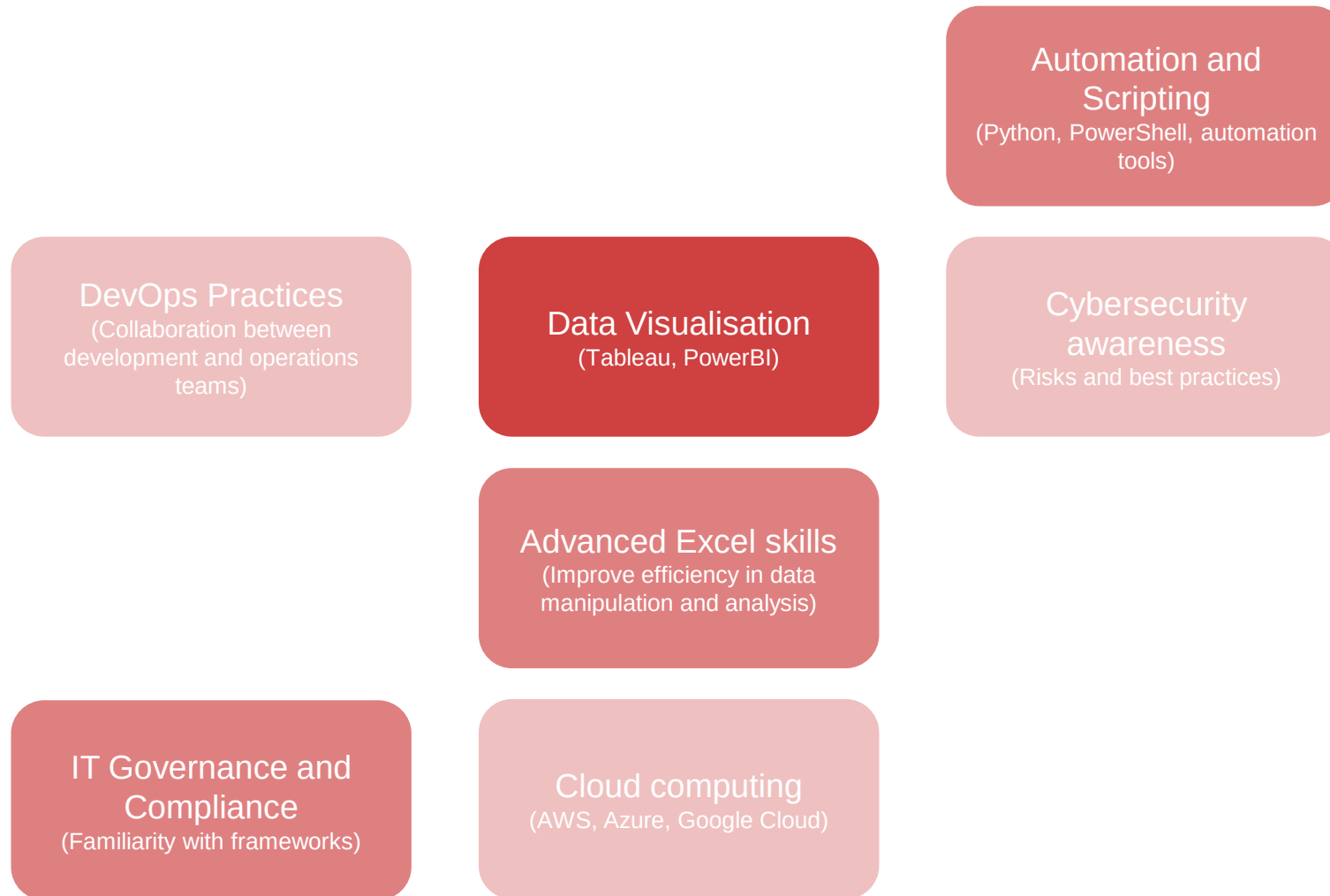
AGENDA



- (IT) Skills evolution for actuaries
- Creative data sourcing and management
- Advanced predictive modelling
- Upgraded coding, operationalization and reporting skills
- Conclusion

Upgraded coding, operationalization and reporting skills

Operationalisation of actuarial/financial computation engines is key for improving data-driven decisions

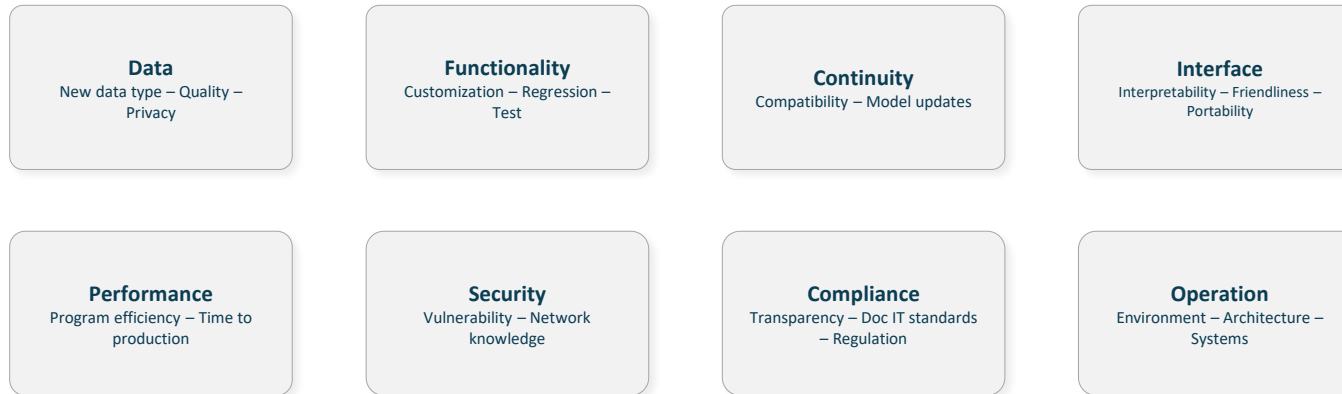


Upgraded coding, operationalization and reporting skills



Pain points with (some) actuarial/financial processes

- Business experts (e.g., actuaries) sometimes **struggle with the efficiency** of actuarial/financial processes considering they must deal with



- It may bring **painful consequences** for companies such as

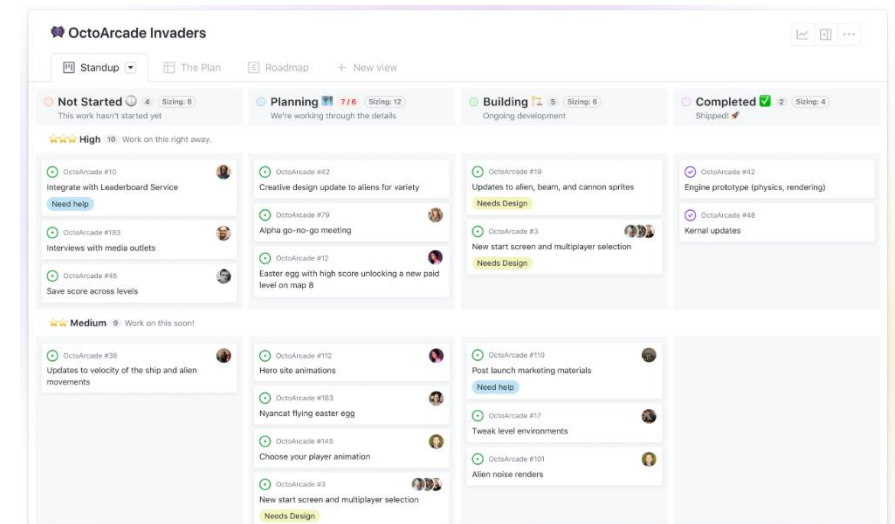
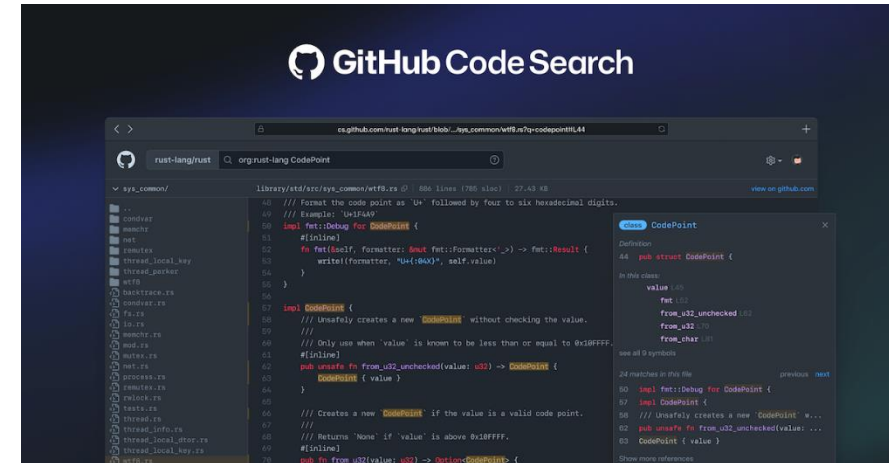


Upgraded coding skills



- Pain points identified
 - **No/bad versioning** of the codes developed
 - Difficult collaboration between developers and organization of tasks difficult
 - Difficulty to push the version in **production**

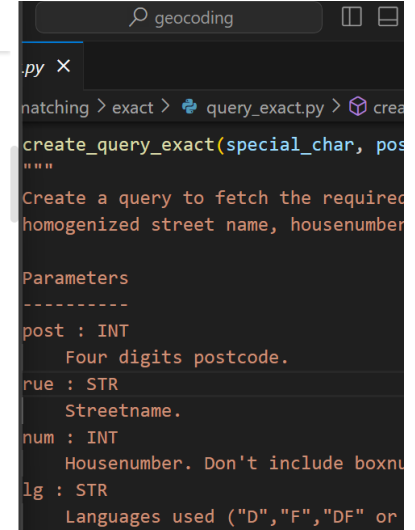
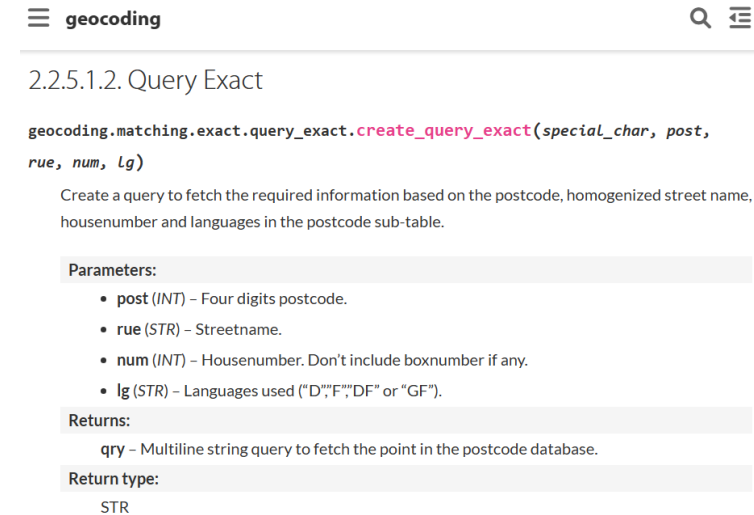
- Some solutions
 - **Git**: free and open-source distributed **version control** system and collaborative coding tool
 - **GitHub/GitLab**
 - Ideal to share and deploy code based on **Git**
 - **CI/CD platform** (Continuous Integration / Continuous Delivery).
 - **GitHub Copilot** is an AI coding assistant allowing suggestions in real-time
 - **GitHub Project** is an online tool which implements the **Kanban** and **Gantt** methodologies in a digital way
 - Other solutions for project management: Trello, Jira, Notion



Upgraded coding skills

Decreasing key people risk with relevant documentation

- Pain points identified
 - Knowledge sharing and **business continuity** are key for actuarial teams
 - **Documentation** is not always up to standards
 - In particular the documentation linked to tools
 - It can create difficulties when changes/improvements are needed
- Some solutions
 - **Code to documentation** tools
 - **Sphinx** is a tool that makes it easy to create intelligent and beautiful documentation (in Python but also in C/C++)
 - Other tools: MkDocs, Doxygen

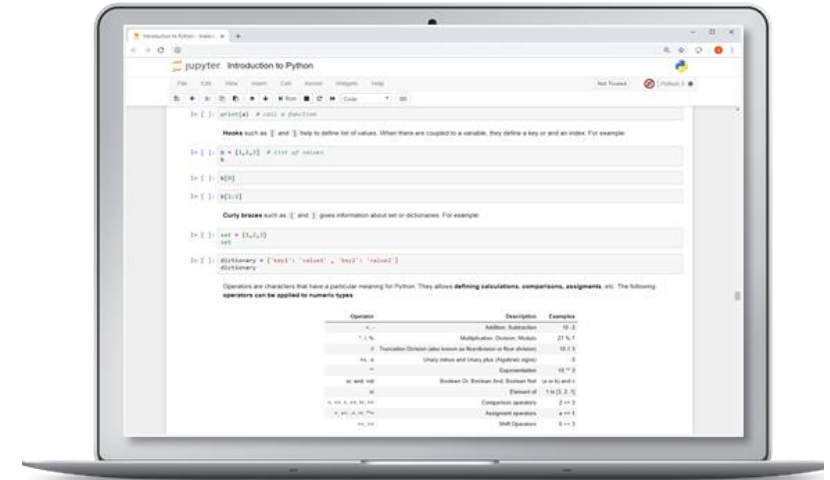


Operationalization and reporting

Developing robust prototypes and coding pipe-lines



- Pain points
 - Developing code in R Studio or another editor does not always allow to create a **clear project pipe-line** incorporating code, results and comments
- Some solutions
 - When prototyping, the cell-based approach of **Jupyter notebooks** is great.
 - They're great for showcasing your work. You can see both the code and the results.
 - It's easy to use other people's work as a starting point. You can run cell by cell to better get an understanding of what the code does.
 - Alternatives: R Markdown, CoCalc

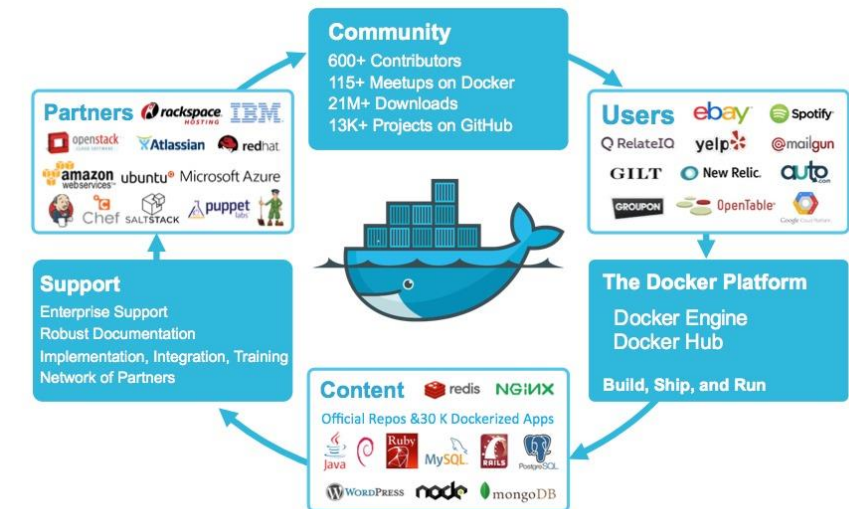


➔ Using notebooks is probably a very good starting point if you want to standardize your coding pipe-line

Operationalization and reporting

Improving durability and efficiency of applications

- Pain points
 - Versions of open-source languages (e.g. R, Python) evolve over time, new packages are developed, or existing packages are adapted → Difficult to **maintain an application over time**
- Some solutions
 - **Docker** is a platform for developers to develop, deploy, and run applications with containers.
 - A container is launched by running an image which is an executable package that includes everything needed to run an application--the code, a runtime, libraries, environment variables, and configuration files
 - Other tools: Containerd, Kubernetes

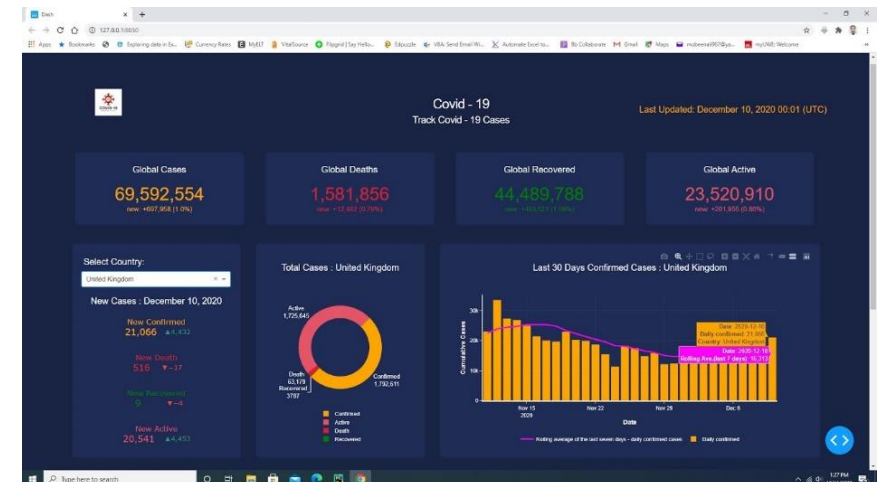
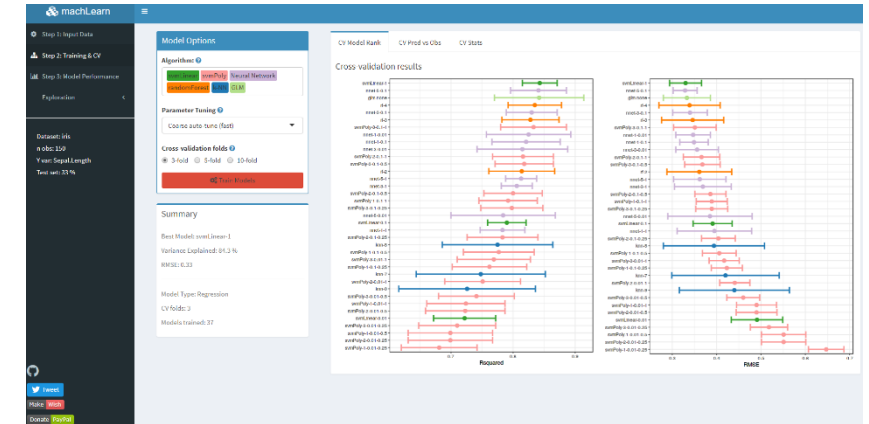


Operationalization and reporting

Visualisation is a must to communicate insight and improve story-telling



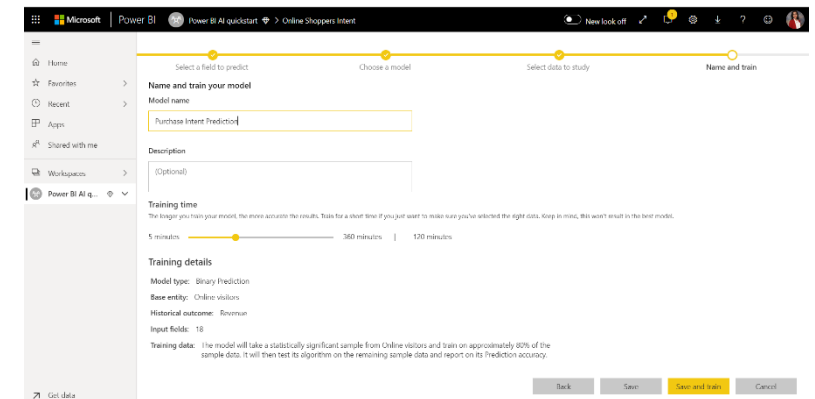
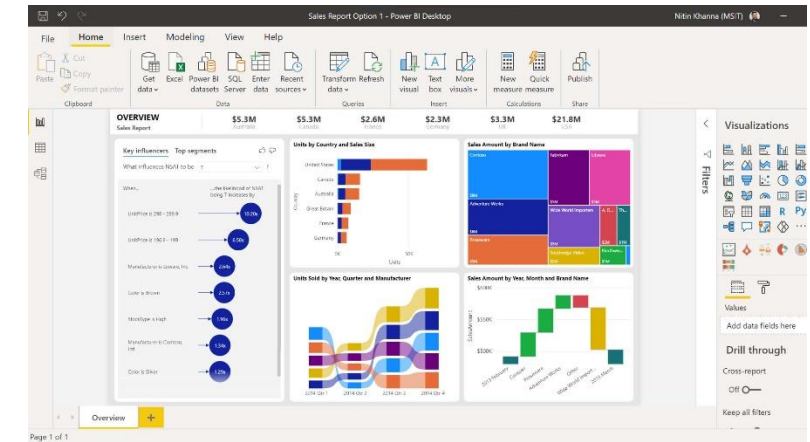
- Pain points
 - **Communicating** results to other stakeholders is complicated as they might not be used to quantitative aspects
 - **Completeness and capacity to drill down** is not available in static reporting
- Some solutions
 - Developing dashboards, i.e. graphical user interfaces containing one or several data visualizations to monitor key performance indicators (KPIs)
 - **R Shiny** is a R package that allows to develop web application using R Studio (code editor)
 - The **shinydashboard** package is also really useful to create dashboards in Shiny.
 - **Dash** is a python framework created by plotly for creating interactive web applications



Operationalization and reporting

Visualisation is a must to communicate insight and improve story-telling

- **Power BI** is a cloud-based business intelligence service suite by Microsoft.
 - It is used to convert raw data into meaningful information by using intuitive visualizations and tables. One can easily analyze data and make business decisions based on it.
 - It offers **drag-and-drop features** and **self-service capabilities**.
- A lot of **other BI tools** exist
 - Tableau, MicroStrategy, SAS BI, Qlik...
- Interesting visualization tools are also available online for **Excel**
 - In addition to the basic data visualization capabilities already embedded through charts, PivotTables or PivotCharts



Operationalization and reporting

Application Programming Interface (API)



- APIs is an entry point to access assets (data, code, services,...) of a provider
- APIs have many advantages
 - Automation and efficiency
 - Integration and simplified access
 - Adaptation and possibility of customization
 - Easiness to share content
 - Secure access to assets
- Examples of use in insurance
 - Digitalisation of operations
 - Increased efficiency in claims management (e.g. upload photos)
 - Developing a partners' ecosystem (aggregators, embedded insurance,...)

Google APIs



AGENDA



- (IT) Skills evolution for actuaries
- Creative data sourcing and management
- Advanced predictive modelling
- Upgraded coding, operationalization and reporting skills
- Conclusion

JoCo2024



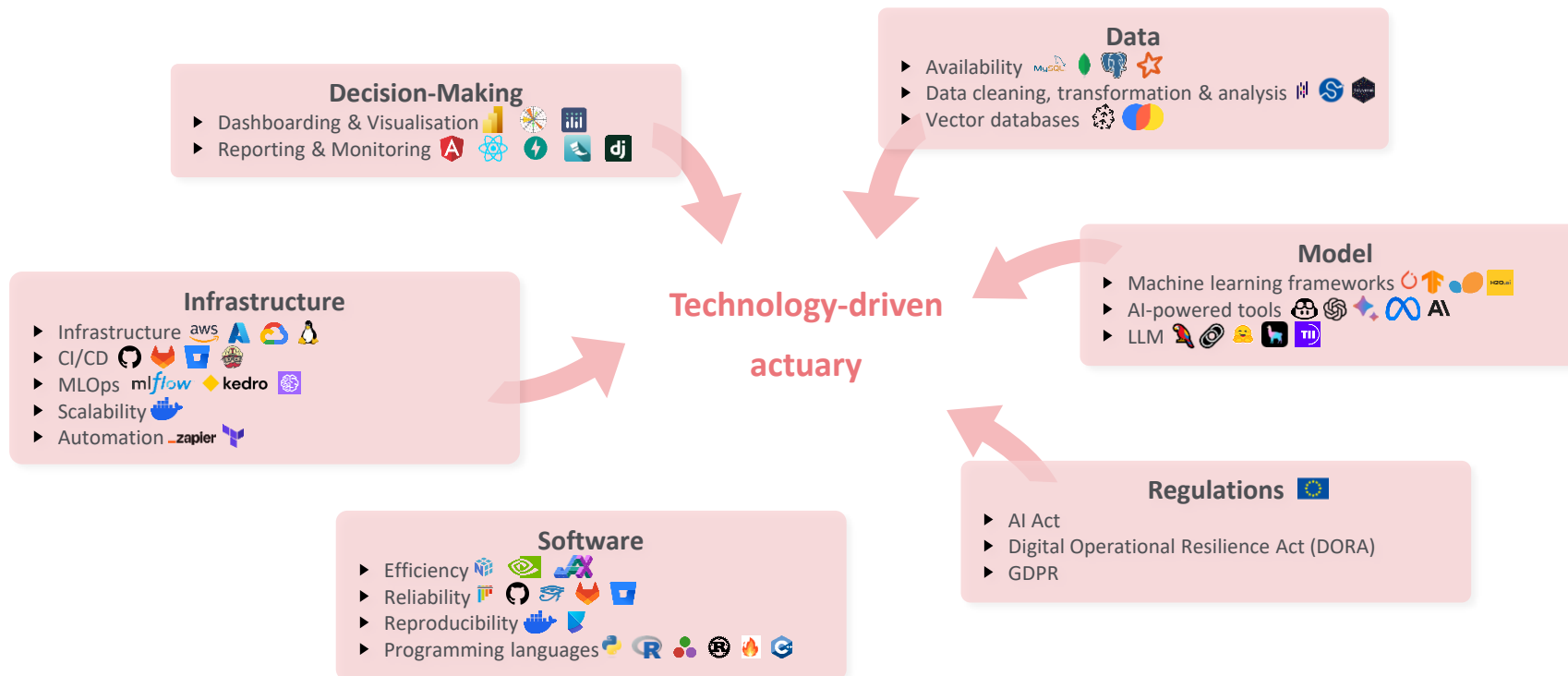
- More than 150 speakers over 3 days on various topics (between 6 and 8 parallel sessions)
- The provisional program is available on <https://www.joco2024.org/program-conference>
- Early-bird initially finished on 31st May but is prolonged until 15th June 2024

The screenshot shows the JOCO 2024 website banner. At the top, there is a navigation bar with links: HOME, PROGRAM, REGISTRATION, PRACTICAL INFORMATION, SPONSORS, and CALL FOR PAPER. The main banner features a large, stylized logo of three interlocking arrows forming a triangle. To the left of the logo is a yellow starburst shape with the text "REGISTRATIONS ARE OPEN!". Below the logo, the text "THE FIRST JOINT COLLOQUIUM OF ALL IAA SECTIONS*" is displayed in large, bold, white letters. Underneath this, in smaller white text, is the asterisked list: "*: ASTIN, AFIR-ERM, IACA, IAALS, IAAHS and PBSS". At the bottom of the banner, the dates "22 to 25 September 2024" and the location "Brussels - Belgium" are written in white.

Conclusion



- Navigating the technological world is not easy as it is **evolving very fast**
- But there is a sweet spot for **actuarial engineers** between business and IT to develop robust actuarial and financial computation engines improving the decision-making process
- Thanks to their **business knowledge**, actuaries will be better placed than any profession to help insurance companies thrive should they **improve their technological skills**



Thank you

Do you have questions ?

Reacfin

Know-How to Risk



Reacfin

Place de l'Université 25
B-1348 Louvain-La-Neuve (Belgium)
T +32 (0) 10 68 86 07
www.reacfin.com

About us

We develop, in partnership with our clients, actuarial & quantitative financial solutions, building on strong data analytics, while securing full transparency and integral knowledge transfer.

 **Xavier Maréchal**
CEO

M +32 497 48 98 48

xavier.marechal@reacfin.com