Institute and Faculty of Actuaries

Enterprise Machine Learning Pipeline

John Ng

12 March 2020

Challenges

- Gartner says 80% of analytics insights will not deliver business outcomes through 2022 and 80% of AI projects will “remain alchemy, run by wizards” through 2020.

  - Gartner (Jan 2019)

- 73.4% of executives report that business adoption of big data and AI initiatives continued to represent a big challenge. 91.5% of firms reporting ongoing investment in AI, but only 14.6% have deployed AI into widespread adoption.

  - NewVantage Partners’ 2020 Big Data and AI Executive Survey

- Data Science initiatives stay at Experimental phase without getting to production?
- How can actuaries leverage data science production process to add value?

Institute and Faculty of Actuaries

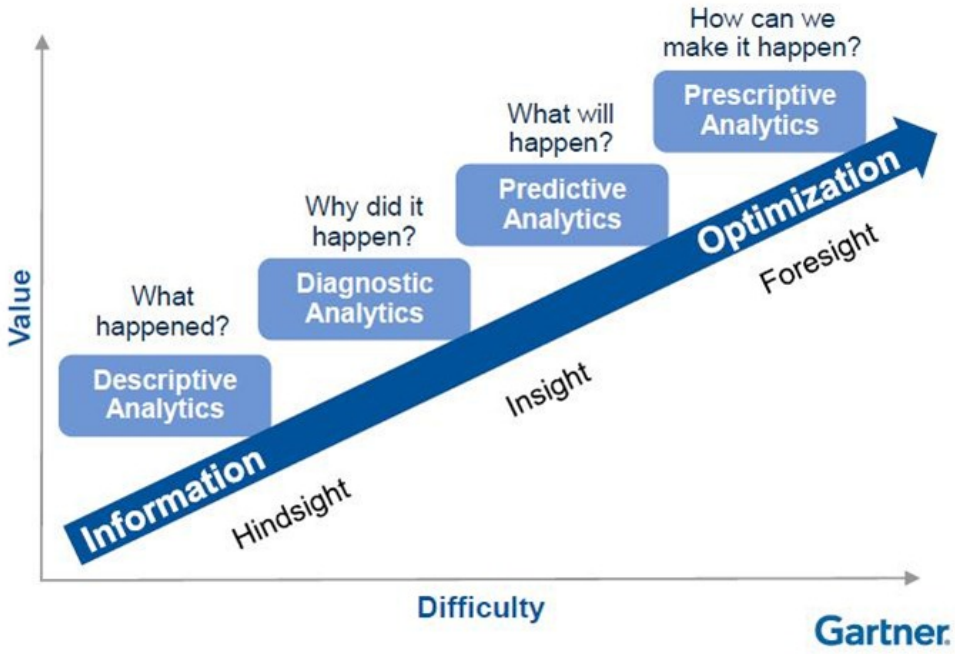
Agenda

- Objective
- Strategy
- AI/ML Pipeline
- Management
- Action

Objectives of Machine Learning Pipeline

- Speed
- Performance
- Integration
- Resilience
- Scalability

Gartner Analytics Ascendancy



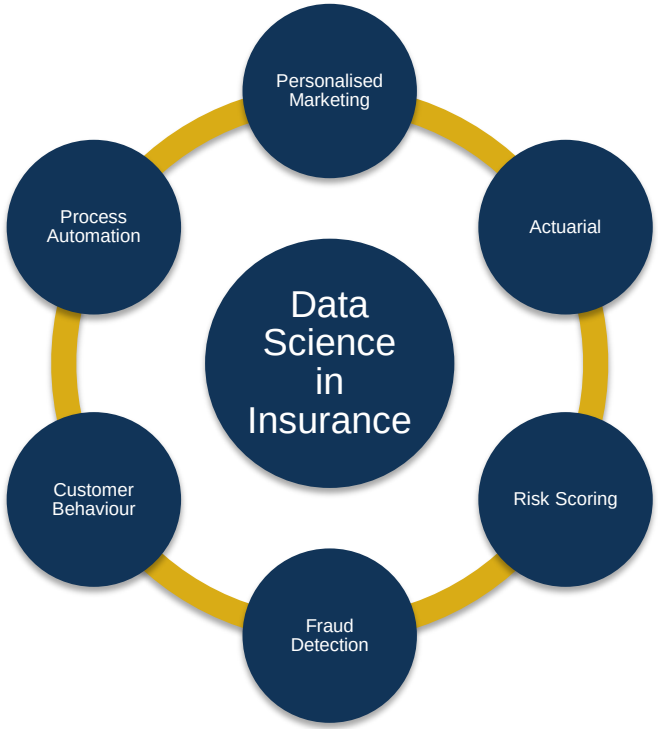
Source: *Gartner (2012)*

Institute and Faculty of Actuaries

Data Science use cases in Insurance

- Call Center Automation
- Chat-bots
- Robo-Advisors
- Paperwork automation

- Conversion
- Persistency/Renewal
- Churn
- Cross-Selling
- Customer Segmentation
- Customer Life-Time-Value (LTV)
- Recommendation Engine



- Pricing Accuracy
- Pricing Sensitivity
- Pricing Optimisation
- Reserving
- Capital Modelling
- Mortality and Morbidity
- Sentiment Analysis

- Claims Prediction
- Risk Granularity
- Automated Underwriting
- Motor Telematics
- Healthcare analytics
- Portfolio Analytics

Institute and Faculty of Actuaries

Institute and Faculty of Actuaries

Enterprise Machine Learning Pipeline

The Architecture

12 March 2020

Institute and Faculty of Actuaries

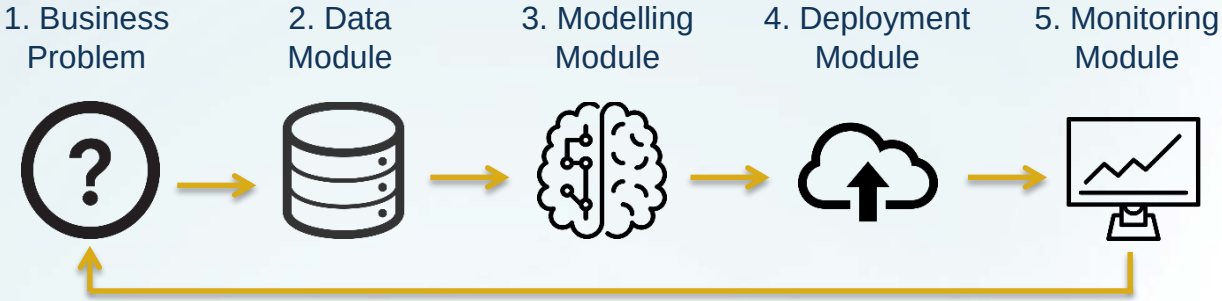
Enterprise Machine Learning Pipeline

1. Business Problem
2. Data Module
3. Modelling Module
4. Deployment Module
5. Monitoring Module

Opportunity | Information | Insight | Execution | Result

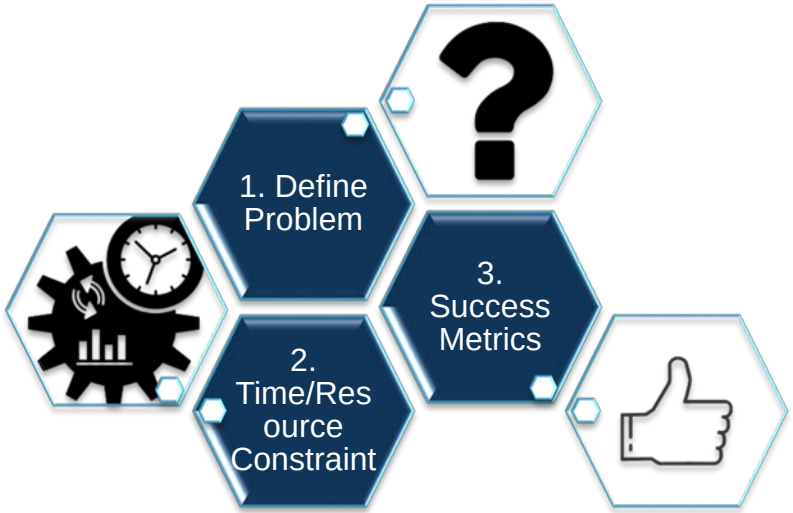
Institute and Faculty of Actuaries

Enterprise Machine Learning Pipeline

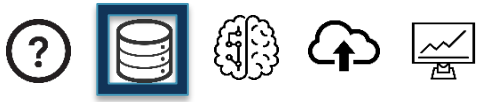


Actuarial Control Cycle

Business Problem



Data Module



Sourcing & Data Engineering

- Data Sources
- Data Connectivity
- Data Engineering
- Data Warehouse

Data Cleaning & Preparation

- Assess Quality
- Formatting
- Outliers
- Missing Data
- Trends

[Auto Reporting]

Data Dictionary

Exploratory Data Analysis (EDA)

- Statistics
- Correlations

[Auto Reporting]
[Auto Visualisation]

Feature Engineering

Feature ..

1. Extraction
2. Transformation
3. Selection

- Expert Driven
- Automatic F.E.

[Auto Reporting]

Feature Store

Data Segregation

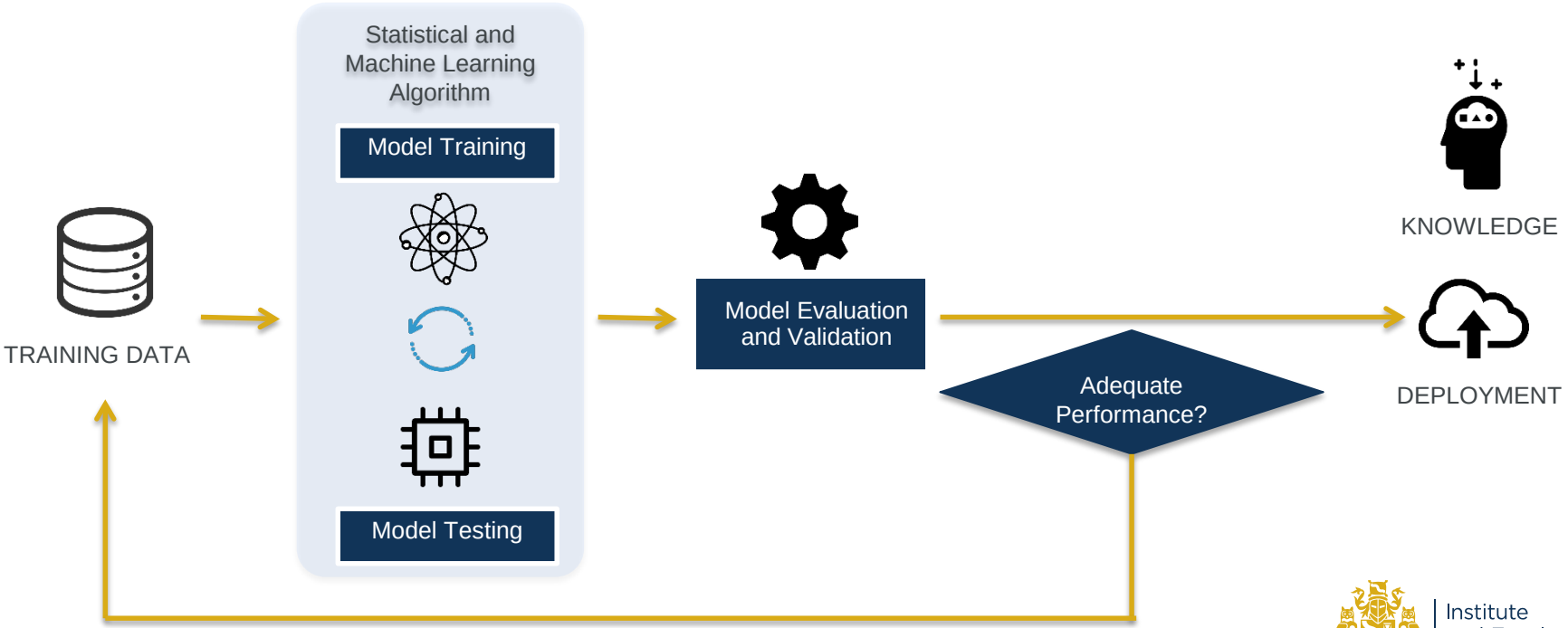
Data into

- Train Set
- Test Set
- Validation Set

[Auto Reporting]

Institute and Faculty of Actuaries

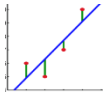
Modelling Module



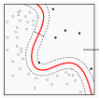
Institute and Faculty of Actuaries

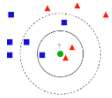
Modelling Module

Model Catalogue | Optimisation Metric | Hyperparameter Tuning

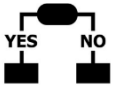

Linear Regression


GLM & Regularization

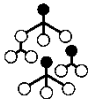

SVM


K-Nearest-Neighbour


Survival Modelling



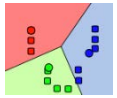
Decision Tree

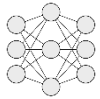

Random Forest


Gradient Boosted Machines (GBM)

XGBoost
Extreme Gradient Boosting


Natural Language Processing (NLP)

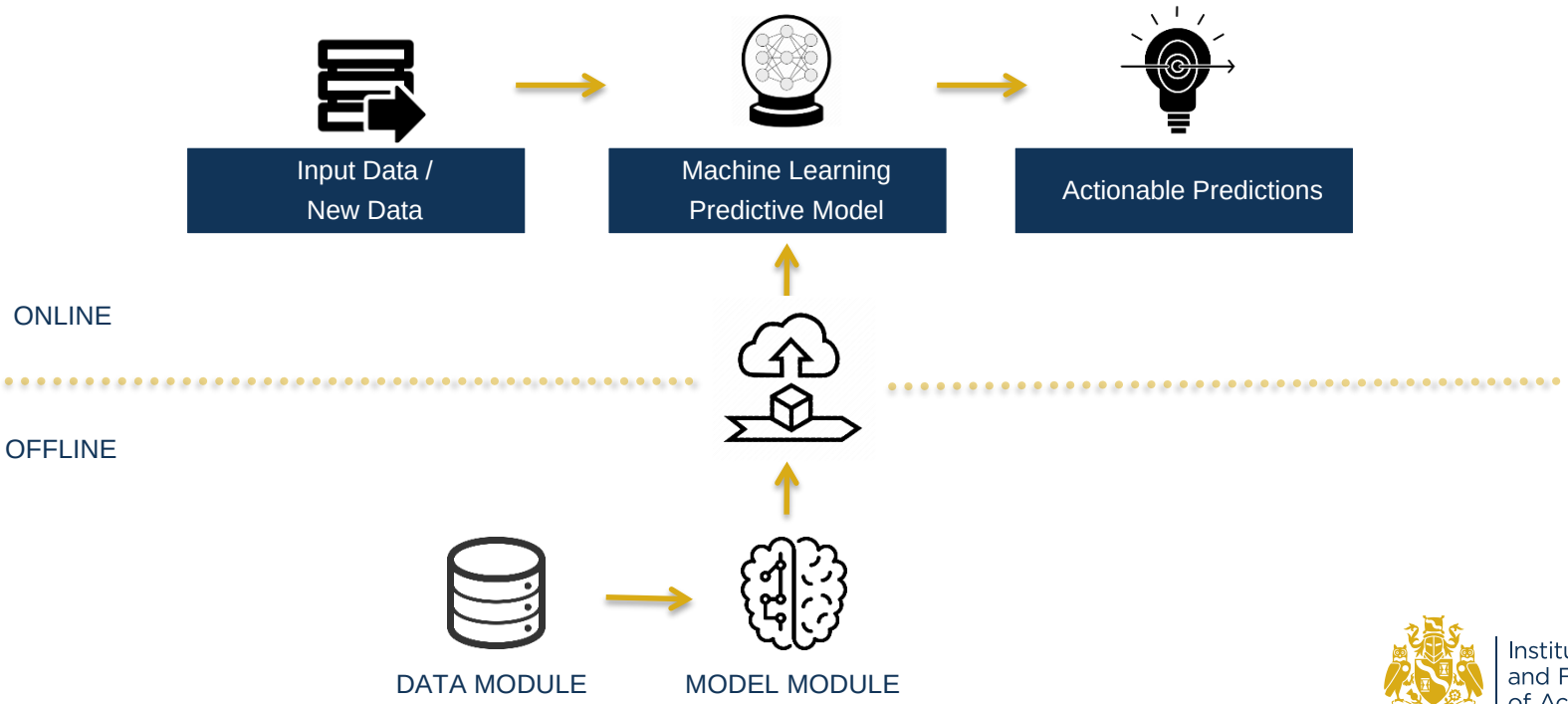

K-means clustering


Artificial Neural Network


Custom Model

Institute and Faculty of Actuaries

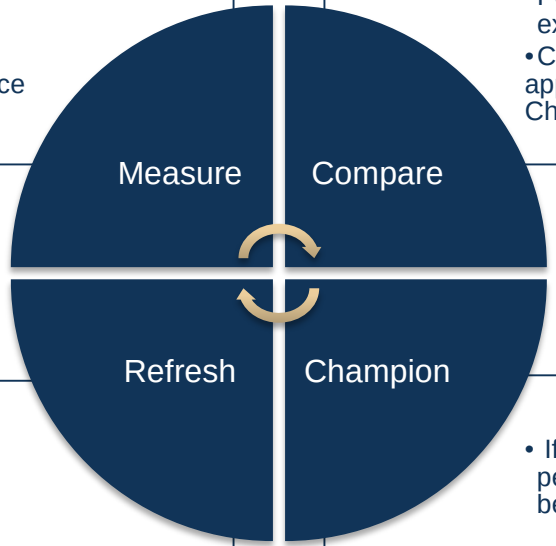
Deployment Module



Institute and Faculty of Actuaries

Monitoring Module

- Constant monitoring of newly deployed model using agreed Performance Metrics
- Actual vs. Expected Performance

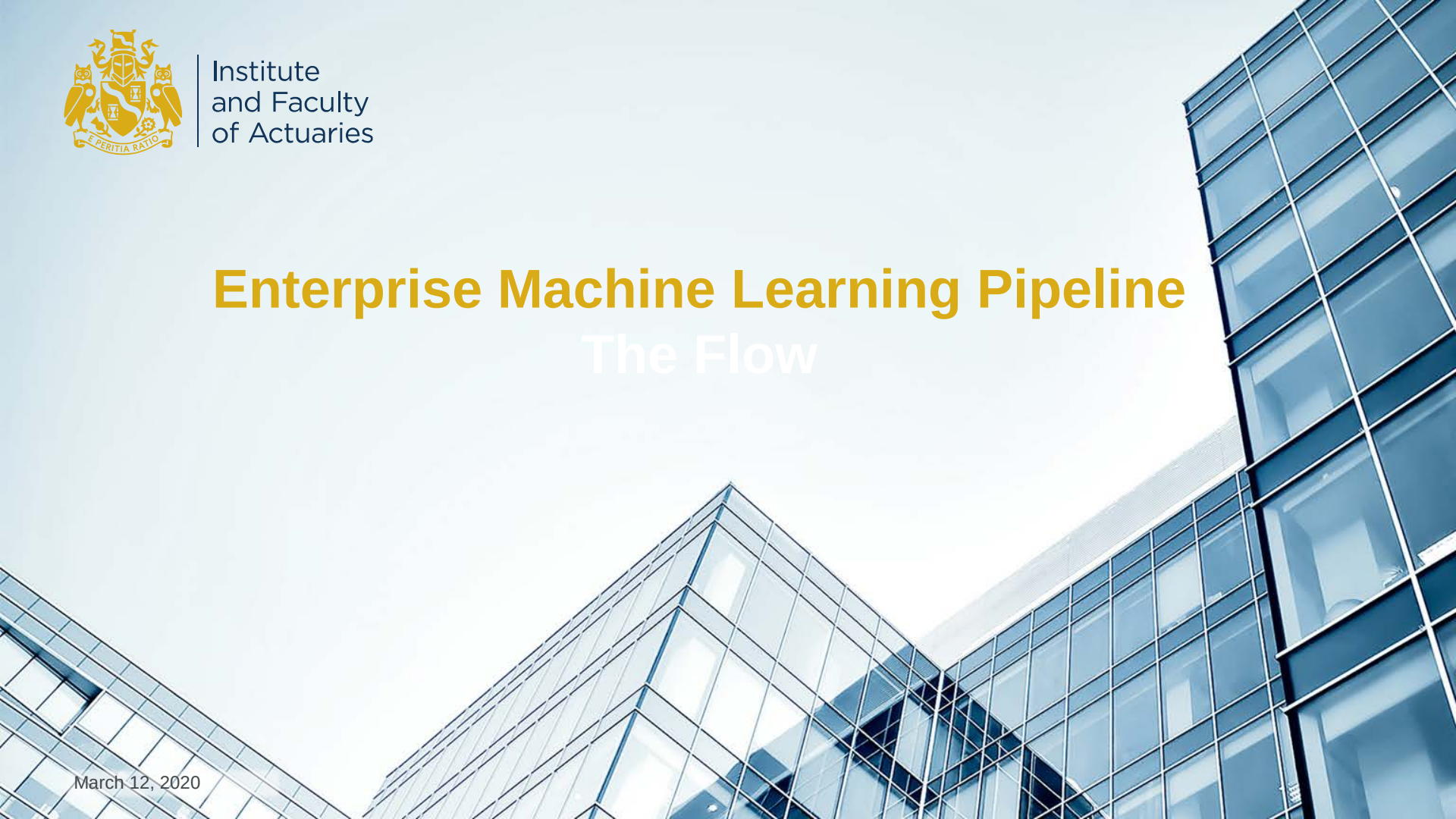


- Perform Champion Challenger experiment (a.k.a. A/B testing)
- Compare against incumbent best approach/ rule/ model (Current Champion)

- Model will eventually degrade (change in data, market etc)
- Rebuild when metric dropped below a determined threshold

- If successful, promote the best performing challenger model to be the new Champion

Institute and Faculty of Actuaries



Institute and Faculty of Actuaries

Enterprise Machine Learning Pipeline

The Flow

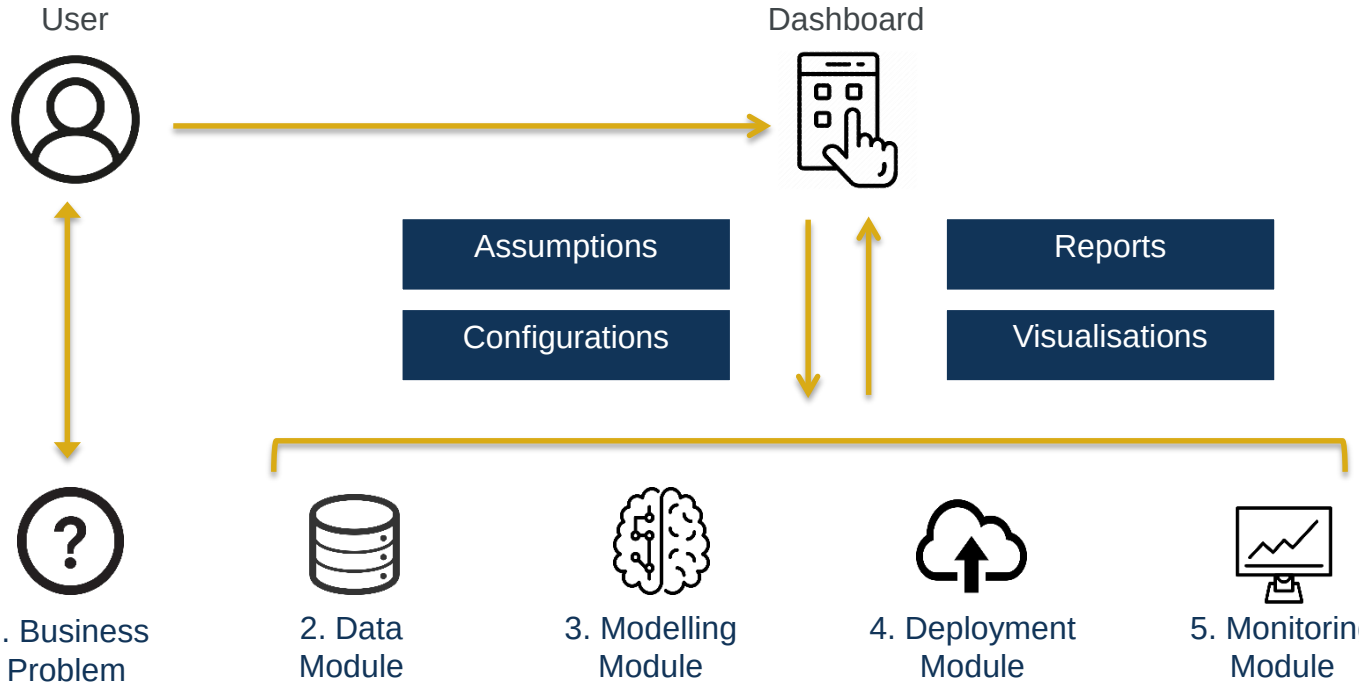
March 12, 2020

Pipeline Automation and Management

- Automation of Processes: Efficiency and Consistency
- Automated Logging, Reporting, Audit Trail
- Error Handling
- Simplify Machine Learning lifecycle development
- Version Control (e.g. GitHub)
- Platform for Business-As-Usual operations, R&D and Proof-Of-Concepts
- Integration into Enterprise
- Scalability

Institute and Faculty of Actuaries

Pipeline Automation: Dashboard



Institute and Faculty of Actuaries

Pipeline Governance

- Iterative Improvement
- Implementation of Best Practices
- Data Protection
- Data Lineage
- Transparency
- Model Explanability
- Access Control and Security

Institute and Faculty of Actuaries

Action

Once having the right team, technology and data:

- Identify opportunities and the right questions
- Define goals and framework
- Build Minimum-Viable-Product that is scalable
- Modelling and Deployment
- Monitoring and Review performance
- Scaling and Maintenance

Actuaries, having business domain knowledge and technical skills, could harness the strength of data science and champion data-driven advancements at organisational level.

Institute and Faculty of Actuaries

Questions

Comments

Institute and Faculty of Actuaries