Institute
and Faculty
of Actuaries

Modular Framework of Machine Learning Pipeline

John Ng MA FIA BPharm

September 14, 2020

Data Science use cases in Insurance

- Chat-bots
- Robo-Advisors
- Customer Service prioritisation
- Paperwork automation
- Unstructured data

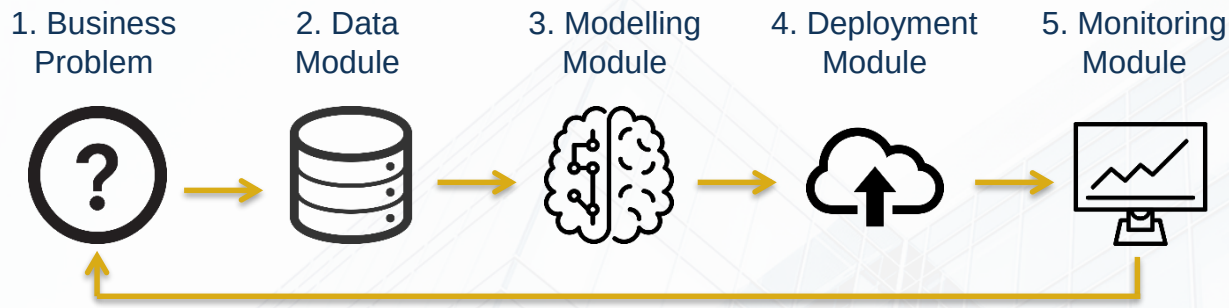
- Conversion
- Persistency / Renewal
- Churn / Lapse
- Cross-Selling
- Customer Segmentation
- Customer Life-Time-Value (LTV)
- Recommendation Engine
- Sentiment Analysis

- Pricing Accuracy
- Pricing Sensitivity & Elasticity
- Pricing Optimisation
- Dynamic Pricing
- Reserving
- Capital Modelling
- Mortality and Morbidity

- Claims management
- Risk Granularity
- Accelerated Underwriting
- Motor Telematics
- Healthcare analytics, Wearables
- Portfolio Analytics

Actuarial Control Cycle

Actuarial Data Science Control Cycle



Data Module






Data Sourcing & Engineering

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

Exploratory Data Analysis (EDA)

- Assess Quality
- Statistics
- Distributions
- Correlations
- Reporting & Visualisation

Data Cleaning & Preparation

- Identify errors
- Formatting
- Outliers
- Remove “post-event” information

Data Dictionary

Feature Engineering

Feature ..

1. Extraction
2. Transformation
3. Selection

- Expert Driven
- Automatic F.E.

Feature Store

Data Segregation

Split into Train, Validation & Test Set

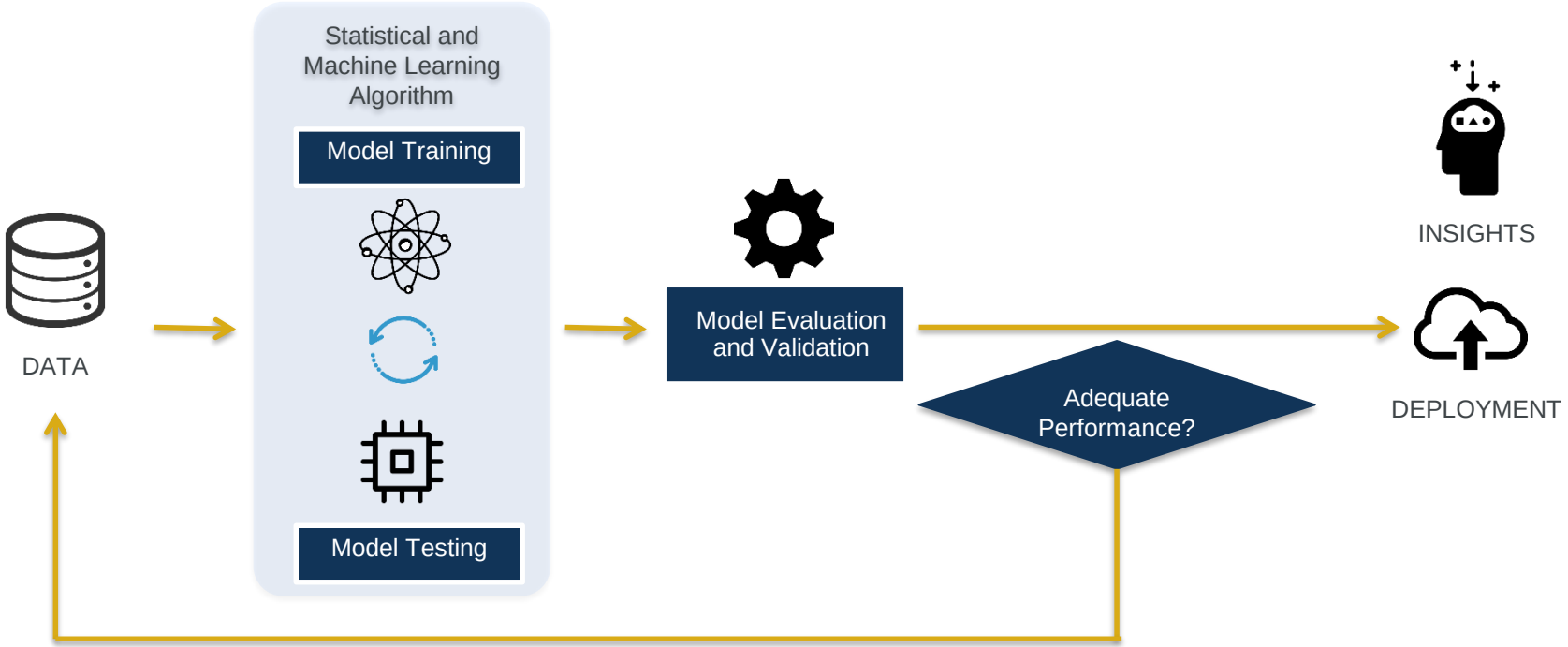
- Random split
- Stratified sampling

Imputation

Impute missing values

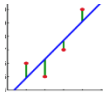
- Fit learner on training set then transform train and test sets
- Fixed imputations (Mean, Median)
- Better approaches: MICE, KNN

Modelling Module

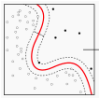


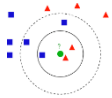
Modelling Module

Model Catalogue | Optimisation Metric | Hyperparameter Tuning

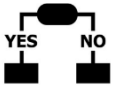

Linear Regression


GLM & Regularization

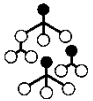

SVM


K-Nearest-Neighbour


Survival Modelling



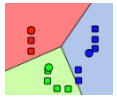
Decision Tree

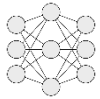

Random Forest


Gradient Boosted Machines (GBM)

XGBoost
EXtreme Gradient Boosting

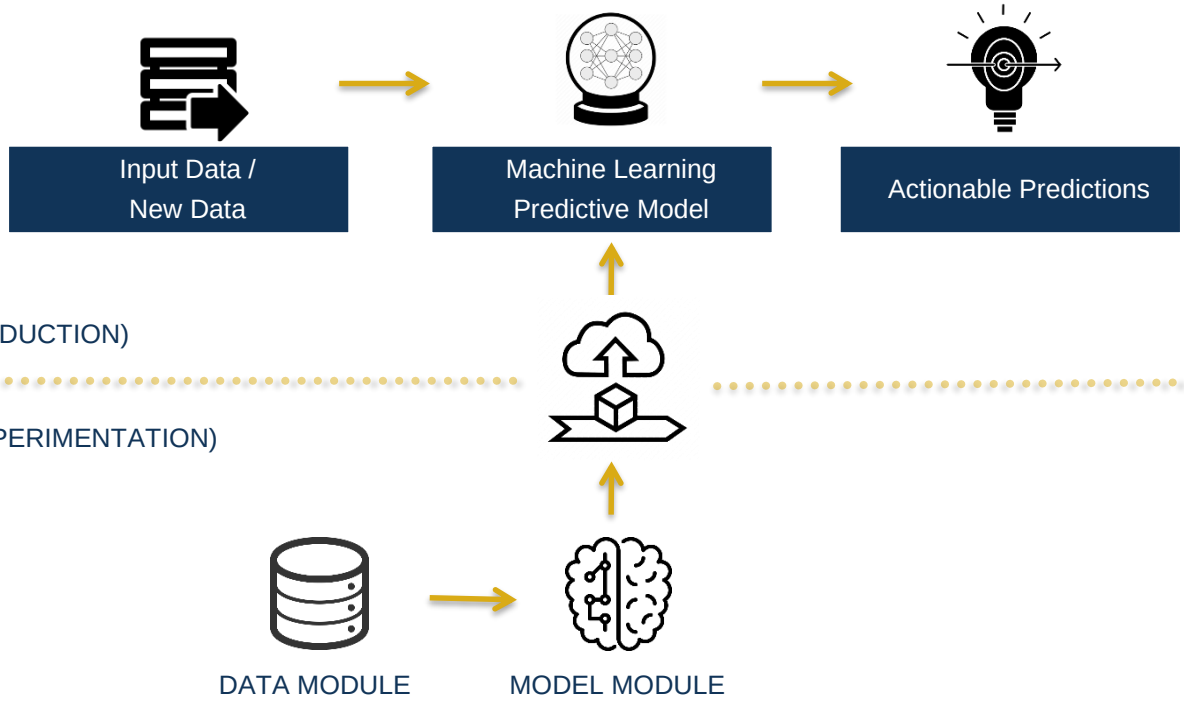

Natural Language Processing (NLP)


K-means clustering


Artificial Neural Network


Custom Model

Deployment Module



Pipeline Operation and Automation

- Speed
- Performance
- Risk Management
- Integration
- Scalability

- Automation of Processes: Efficiency and Consistency
- Simplify Machine Learning lifecycle development
- Best-in-class algorithms for better prediction accuracy
- Leverage best practices in data across enterprise
- Automated Logging, Reporting, Audit Trail
- Error Handling
- Integration into Enterprise
- Common Platform for Business-As-Usual, R&D and Proof-Of-Concepts
- Version Control (e.g. Git)
- Scalability & Iterative Improvement

Pipeline Governance

- Ethics, Fairness
- Regulatory requirements
- Data Protection
- Data Lineage
- Model Explainability / Explainable AI (XAI)
  - SHAP, LIME, DeepLIFT, permutation feature importance
- Access Control and Security

Five Models of Pricing Operation

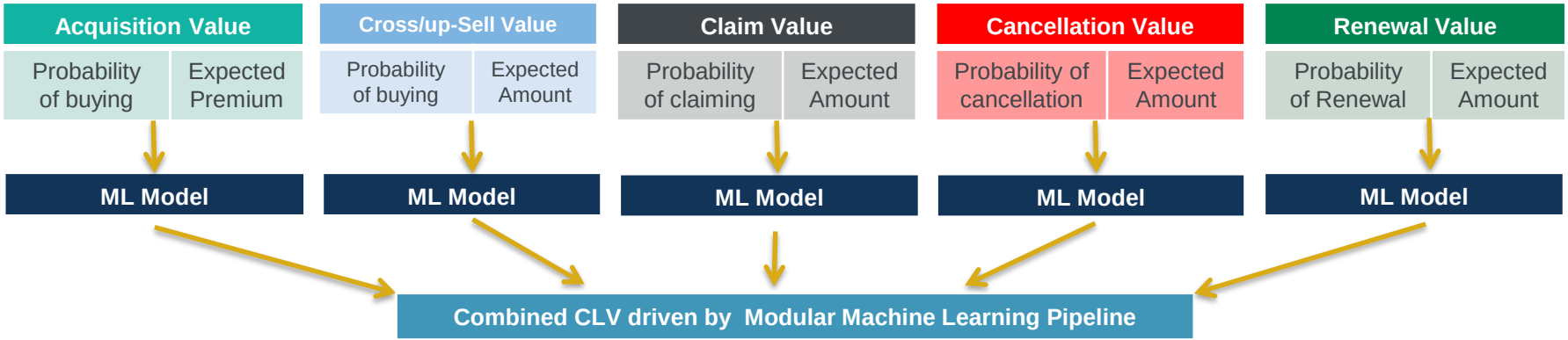
Tariff	Qualitative	Cost Plus	Distribution	Industrial
• Regulator has significant influence over the rates	• "Correct" pricing cannot be determined purely by numerical analysis and subjective factors play a significant role • Data maybe incomplete or not exist	• Statistically driven analysis • Based on **expected cost of claims**, appropriately loaded for expenses, profit etc • Typically single distribution channel	• Price also allows for non cost elements such as **propensity to shop around, price elasticity** • Pricing strategy for similar products being managed across multiple distribution channels	• Typically domain of very large insurers • multiple brands, channels, countries • **Machine oriented approach** • Focus on operating efficiency and economies of scale

where Machine Learning Pipeline can add value

Source: GRIP report

Application 3: Customer Lifetime Value (CLV)

- Definition: *The net present value of a customer during entire relationship with the company*
- Customer Lifetime Value = Present value + Future Value
  - Present value = Premiums + cross/up-sell revenue – Claim costs – Activity-based costs (ABC)
  - Future value = (Premiums + cross/up-sell revenue – Claim costs – Activity-based costs (ABC) – Cancellation)/$(1+i)^t$



Application 3: Customer Lifetime Value Segmentation

CLV ML pipeline helps you to make smart decisions (decision science) and grow business

New Customers: Acquisition Lifetime Value

- Pricing
- Inform marketing target profiles
- Generate sales leads for new customers + prioritisation
- Manage customer service resources
- Cross sell and up sell
- Personalised products
- Product designs or features
- Channel optimisation (affinity partners, price comparison websites)

Existing Customers: Future Lifetime Value

High Value customers

- Cross sell and up sell
- Reduce churn and improve persistency
- Personalised servicing
- Selective discounting and offers

Low Value customers

- Termination or reduce cost of service

Application 3: Customer Lifetime Value Optimisation

CLV Machine Learning Pipeline

Optimiser

Commercial Price associated with
Optimal portfolio CLV value
(subject to constraints)

- Pricing Optimisation
- Granular Price Elasticity
- Dynamic Pricing