Society of Actuaries in Ireland

Good Practices in the Application of Predictive Analytics

Eamonn Phelan FSAI, FIA, CERA
Eileen Burns FSA, MAAA

Disclaimer

The views expressed in this presentation are those of the presenter(s) and not necessarily those of the Society of Actuaries in Ireland or their employers.

Agenda

- Applications of predictive analytics
- SOA research objectives
- Findings from SOA survey
- Good practices for predictive modelling
- Case Study
- Q&A

Data Science and Predictive Analytics

Machine Learning

Artificial Intelligence

Data Collection

Data Cleaning

Data Analytics

Predictive Modelling

"The process of developing a mathematical tool or model that generates an accurate prediction"
– Max Kuhn

Resourcing and governance

Data Strategy

Business Intelligence

Applications of Predictive Analytics

Cross Selling and Discounts

- Offering discounts for purchasing multiple product types

Data Validation and Imputation

- Identifying unexpected data patterns & dealing with missing data

Quotations and Pricing

- Deriving better rating factors and asking fewer u/w questions

Model Validation

- Forecasting future exposure in internal model

Customer Behaviour

- Identifying key drivers of option take-up, fund switches, lapses, etc.

Inforce Management

- Creating behavioural profiles for distinct customer segments

Programming languages

R

python

MATLAB

julia

sas

Data Science Tools

Decision Tree

Linear Regression

GLM

Random Forest

Gradient Boosting

Neural Networks

Simpler → More advanced

SOA Research Objectives

Survey: Capture leading practices among industry participants

Literature review: Research sources internal and external to life insurance

Considerations: Distill findings into areas to specifically address

Case study: Demonstrate leading practices via a case study

Survey Design

- SOA Survey
  - Predictive Analytics
  - Approx. 150 Responses from SOA members
  - Focused on
    - Business applications
    - Data acquisition and preparation
    - Algorithm and software selection
    - Model evaluation, implementation and governance

- Milliman Irish Client Survey
  - Data Science
  - 22 Irish-based life and health insurers and reinsurers
  - Focused on
    - Overall data science strategy
    - Data collection
    - Process and technical application
    - Resourcing and governance
    - Benefits and challenges

Key Findings

- SOA Survey
  - No standardised approaches to applying predictive analytics techniques
  - Wide variety of applications of predictive analytics
  - Business/domain knowledge is very important at a number of stages
  - Simplicity and transparency are key determinants in algorithm selection
  - R is the leading language used
  - Work to do on model governance
- Milliman Irish Client Survey
  - Over 75% expect to be using data science within the next 3 years, with over 35% already making it a point of focus
  - Most common uses of data science right now involve either assessment of customer behaviour or assumption setting
  - Biggest challenges facing companies involve a lack of infrastructure and technology, cyber risks, regulatory expectations, a shortage of talent, data quality, and access to data

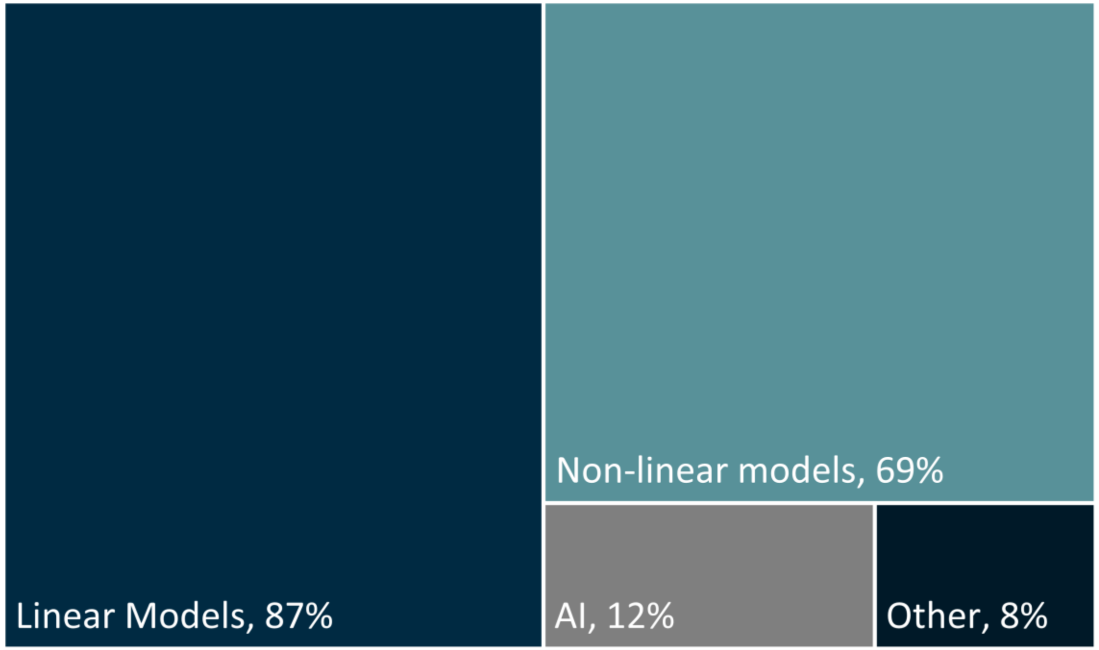
Business Applications

AREA	SOA STUDY		IRISH SURVEY	
Actuarial	Pricing	51%	Actuarial assumption setting	32%
Risk	Underwriting	33%	Monitoring for fraud	27%
Operations	Claims	32%	Optimising operational processes	27%
Customer Service	In-force management	24%	Understanding customer behaviour	41%
Compliance	Compliance	0%	Ensuring compliance standards	14%
Other	Other	8%	Other	14%

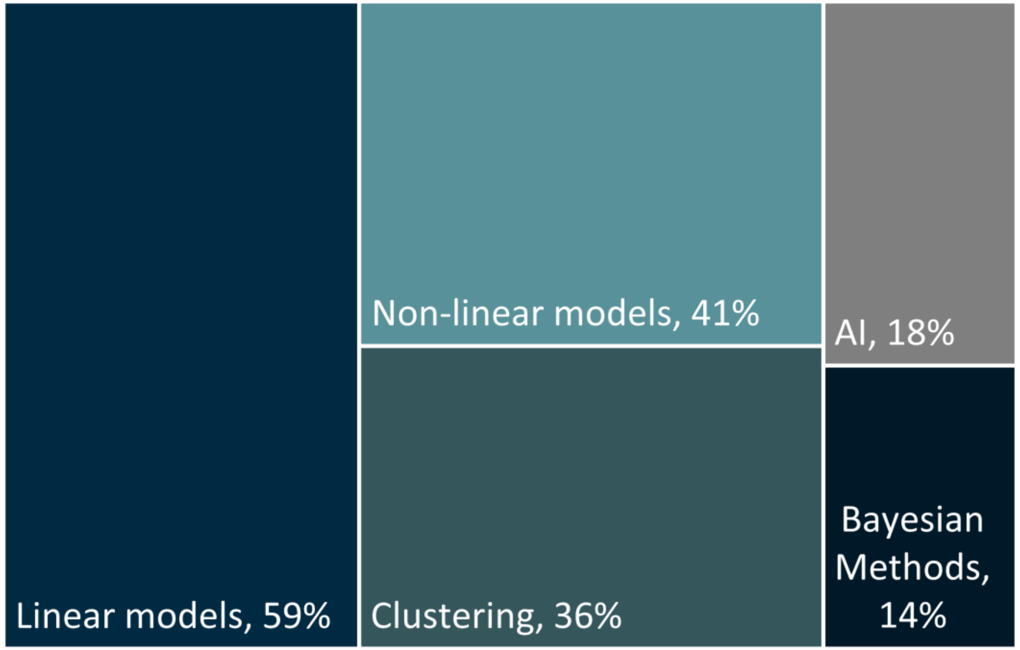
- Milliman – For which business decisions or applications is Data Science used at your company?
- SOA – To which of the following business areas have you applied predictive modelling?

Techniques

- SOA – which technique(s) do you use for predictive modelling?

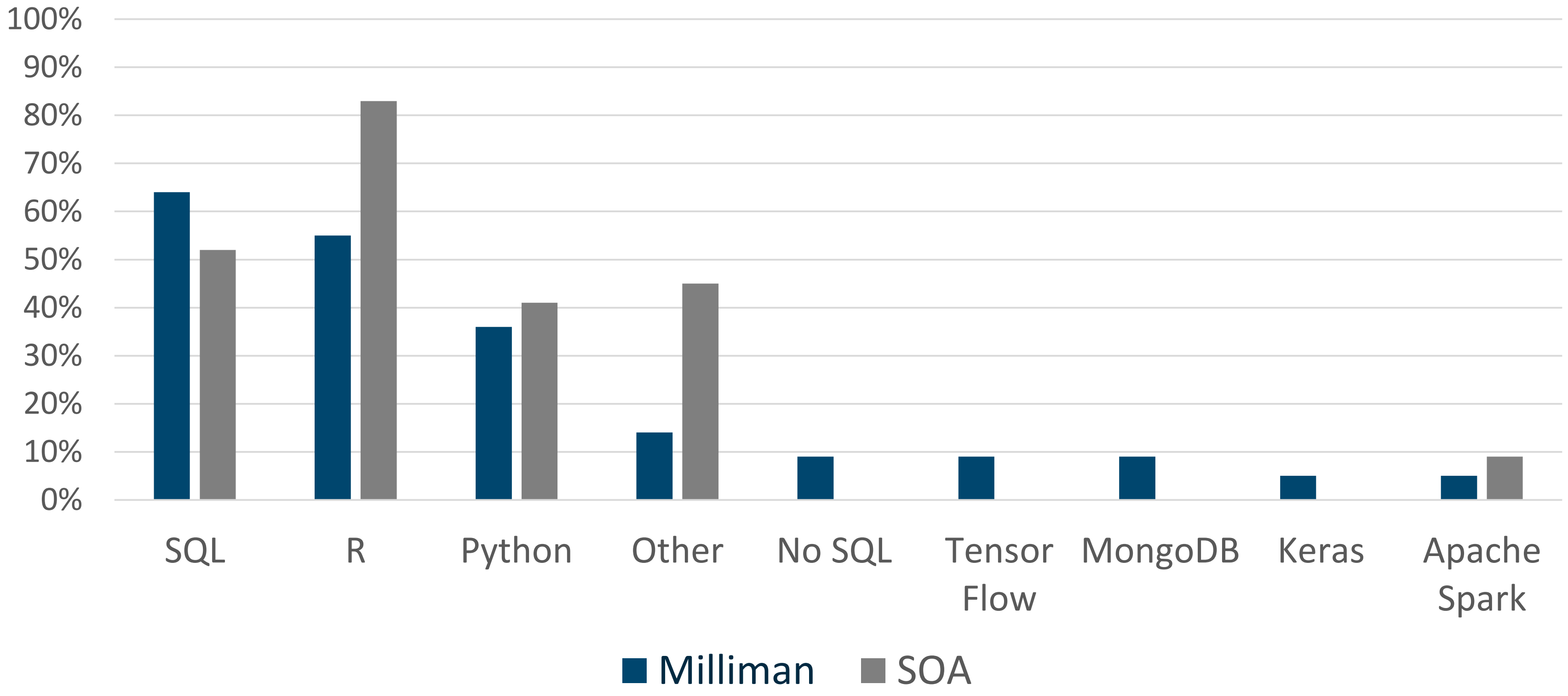


- Milliman – which of the following types of tools or techniques have you used in the application of Data Science (or plan to use in the next 3 years)?



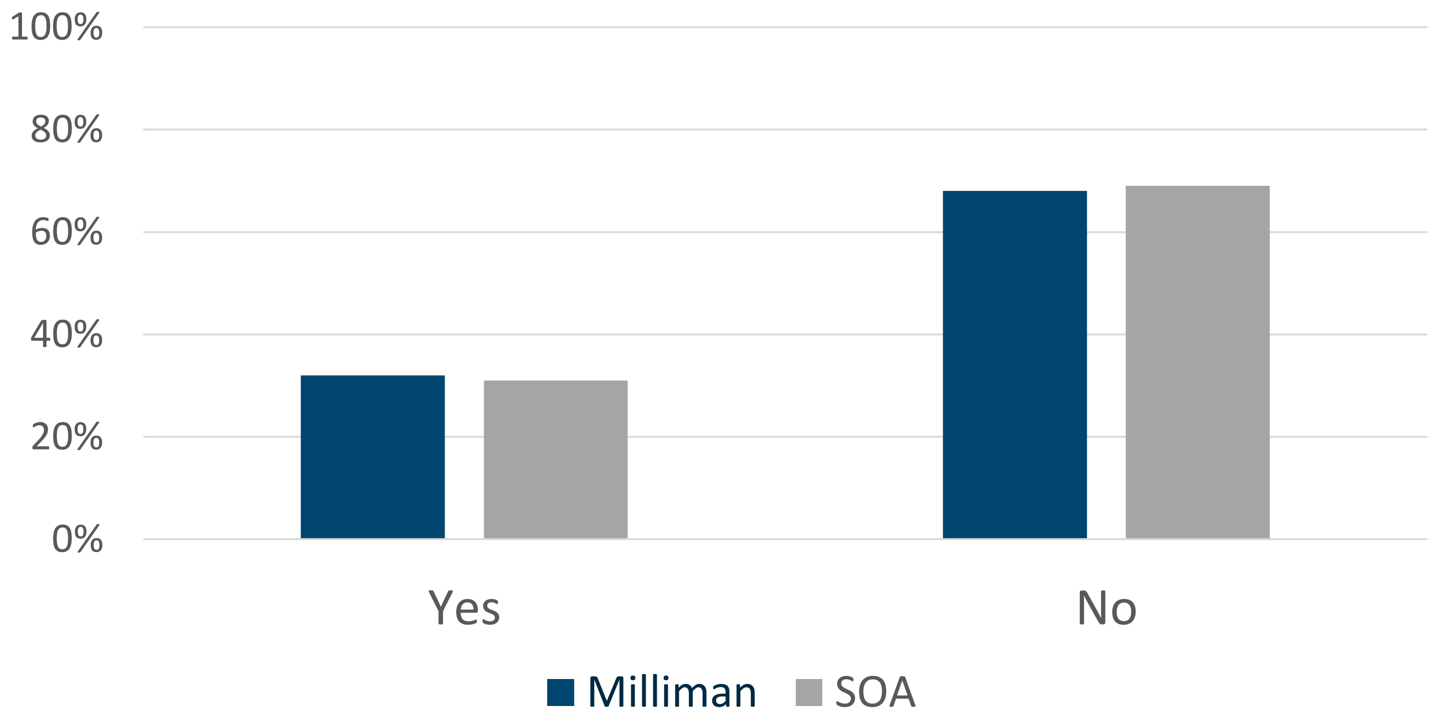
Software Selection

- SOA – What software/language(s) do you use for predictive modelling?
- Milliman – Which of the following types of tools or techniques have you used in the application of Data Science (or plan to use in the next 3 years)?



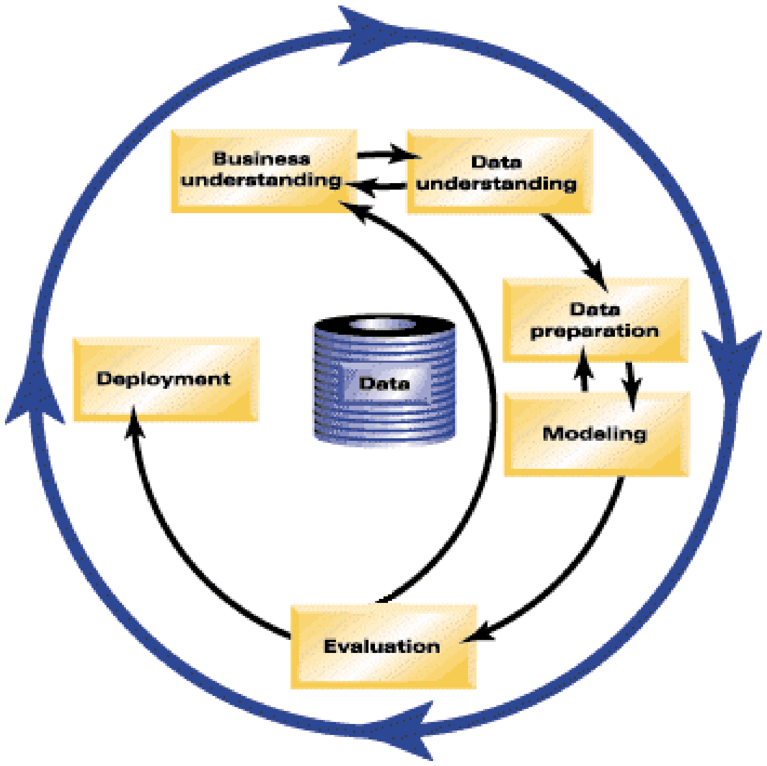
Governance

- SOA – Does your company have a modelling governance framework for your predictive modelling work?
- Milliman – Does your organisation have internal standards governing the use of data science?



Areas of focus

- Project objective
- Data acquisition and preparation
- Algorithm selection
- Feature engineering and selection
- Model evaluation and measurement
- Model deployment
- Model governance
- Software selection



Cross-Industry Standard Process for Data Mining (CRISP-DM)

Project objective

Project objective

- **What are the primary business objectives?**
- How will this provide value to the business and to the customers?
- What are the specific modeling objectives?
- **What is the organizational context?**
- What data will be needed?
- Are there existing limitations to accessing and using the data?
- What resources are available?
- What is the contingency plan for unexpected delays?
- How do we define and measure the model's success?
- What is the end state?

Data acquisition and preparation

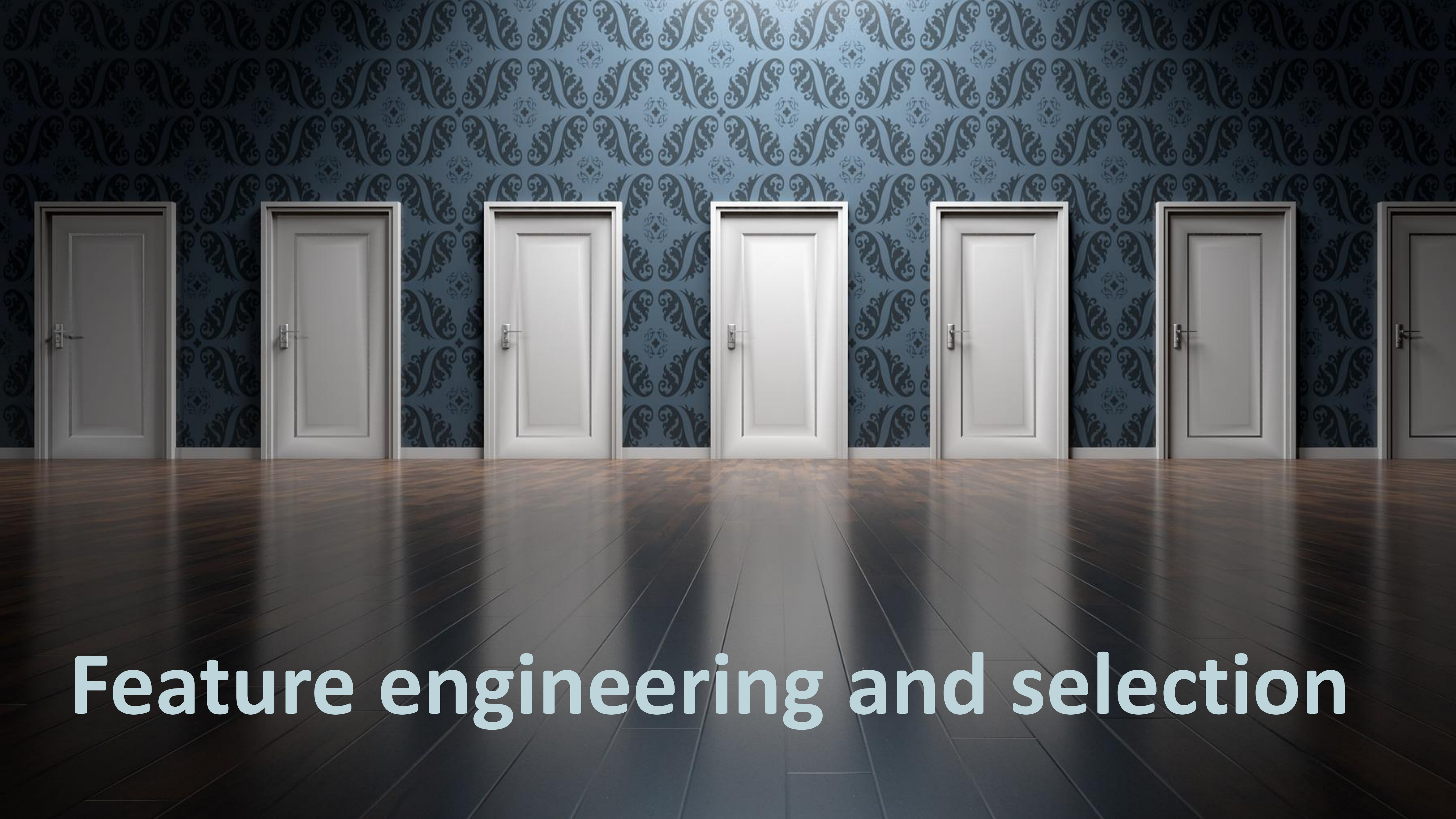
Data acquisition and preparation

- **What data is available internally and externally?**
- Who will gather the data? Depending on the breadth of data sources, this may be one for several people.
- Where will the data be stored once acquired and how will it be accessed?
- What issues and challenges are anticipated, and what is the plan for addressing them?
- What checks can be automated?
- **What will happen if the person familiar with the data leaves the company or the team?**
- What will happen if the data warehouse or underlying data source changes?
- How will the team handle data security, HIPAA compliance, General Data Protection Regulator (GDPR) and other confidentiality requirements?

Algorithm selection

Algorithm selection

- What is your methodology for selecting an algorithm?
- What are the pros and cons of your candidate and selected algorithms?
- Does your selection fit within requirements of stakeholders?
- **Will your chosen algorithm allow you to maximize predictive accuracy relative to requirements for interpretability?**
- **How will you identify and document limitations? How will you effectively communicate them to other stakeholders and model users?**



Feature engineering and selection

Feature engineering and selection

- What is your plan for feature engineering?
- **What steps can you take before feature selection to ensure a smooth process?**
- What is your plan for addressing collinearity of variables?
- **What method will you use for feature selection?**
- What limitations are implied by regulatory, legal, or privacy considerations?
- How will your model evaluation plans affect the preparation of your modeling data?
- How will you verify that the data support your assumptions about the underlying relationships between features and the response, e.g., linearity?
- How will you identify and document limitations?

Model evaluation and measures of success

Model evaluation and measures of success

- How will you ensure you do not overfit the model, balancing the bias-variance trade-off?
- **What metrics will you use to evaluate the relative performance among your candidate models?**
- What visualizations will you use to evaluate the relative performance among your candidate models?
- Will you address credibility of the model predictions, and if so, what measures of credibility will you use?
- Do the training and testing data sets reflect a diversity of scenarios?
- What other considerations will help you choose between models?
- How will you measure the performance of the model in production?
- What are your business measures of success?
- **How will you communicate the value and limitations of your model to all stakeholders?**

Model deployment

Model deployment

Implementation

- How will you document the model and associated assumptions to communicate to users?
- Will all the data required by the model be available once it is deployed?
- **How will the model be operationalized?**

Validation

- How will you check that the model is performing as expected? How frequently will these checks be done?

Updates

- How often or under what circumstances will you retrain the model on updated experience?
- How will you recognize and handle new data that implies the conditions under which you originally fit the model are changing?
- **If you find an error in your modeling process, how will you implement a fix to the model in production?**

Model governance

Model governance

- Will your predictive models be recorded in your company's model inventory?
- **How does your predictive model fit into your company's model governance policy?**
- How will you work with IT on data governance and establish a balance of responsibility and information sharing?
- How will you approach version control to ensure that no unintended changes make their way to the end user?
- **If you are just getting started, what is a minimum requirement for model governance based on the risk level of the model, and what is your plan for improving governance in the future?**
- Who will be responsible for each portion of model governance?
- How will you audit compliance with your model governance policy and procedures?
- How will the risks and/or limitations of your model be communicated?

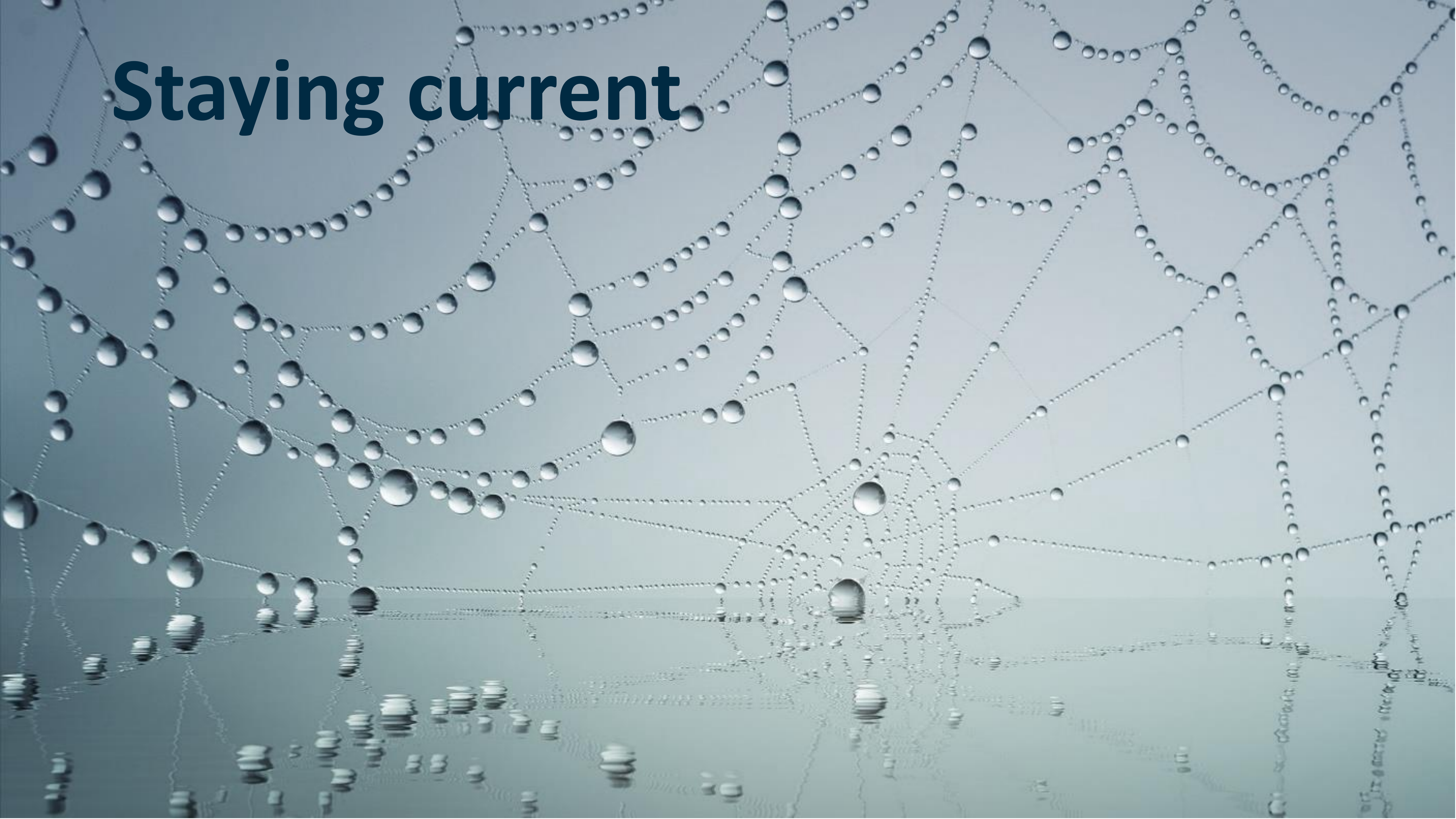
Software selection

Software selection

- **Core competency: Do you want to build your software solution, or do you prefer to purchase an application or framework that does not require programming?**
- Cost: What is your budget, and what is the scope of the project?
- Analytics: What predictive modeling capabilities will you need?
- Visualization: How do you want to deliver your data, models, and results? Who will want to view them? How interactive does the delivery mechanism need to be?

- Other: Processing speed and distributed computing, cross-functionality, client or stakeholder standards, availability of support for software, **existing knowledge within company or modeling team, ability to hire additional qualified practitioners**

Staying current

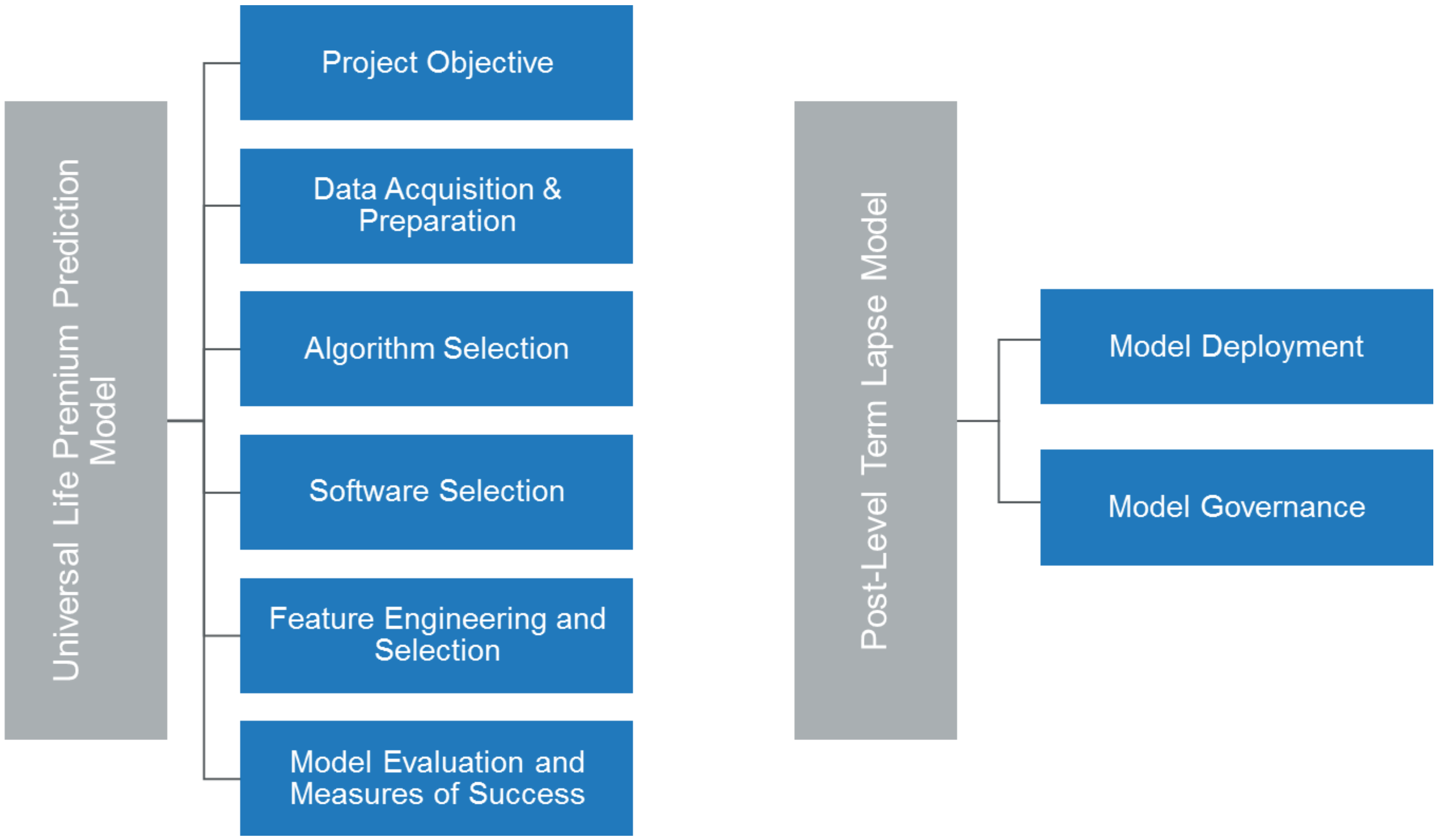


Staying Current

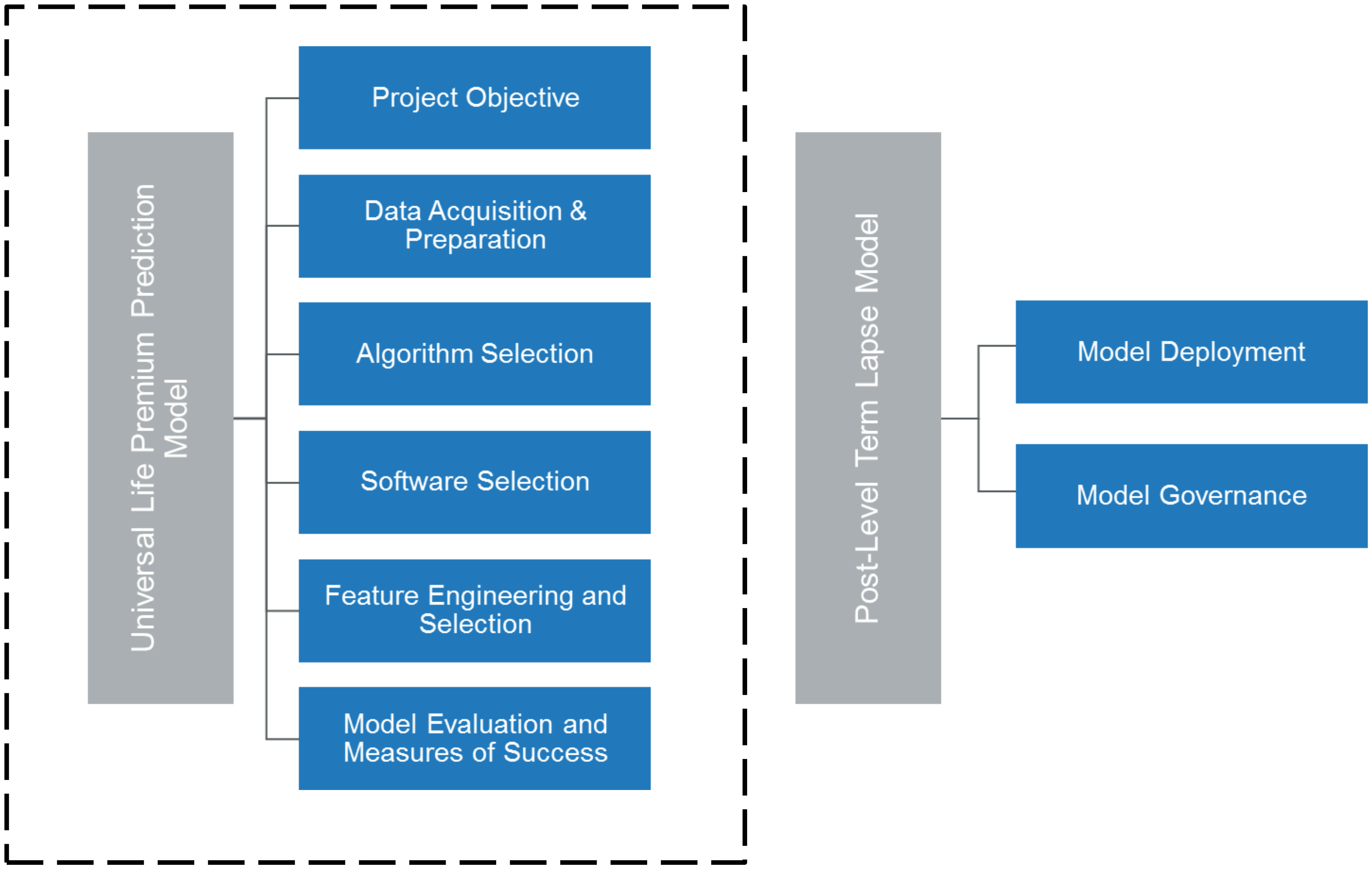
- Web resources
- Social media
- Community
- SOA Predictive Analytics and Futurism Section!

Case Study

Case study structure



Case study structure



Project objective

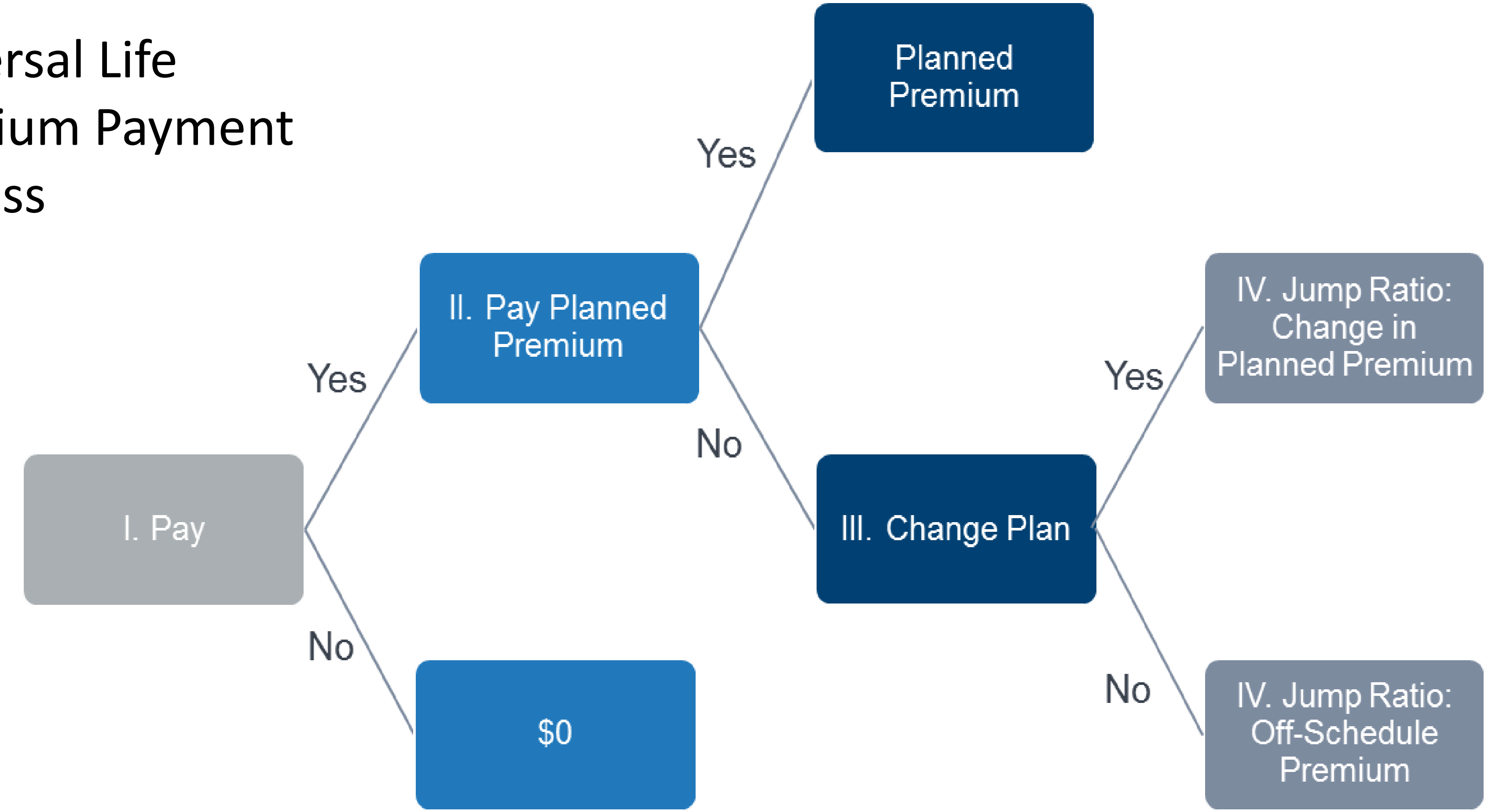
- Predict a policyholder's premium payment amount for the next month on a universal life (UL) policy
- Eventually, predict monthly payment patterns several years in advance

Data acquisition and preparation

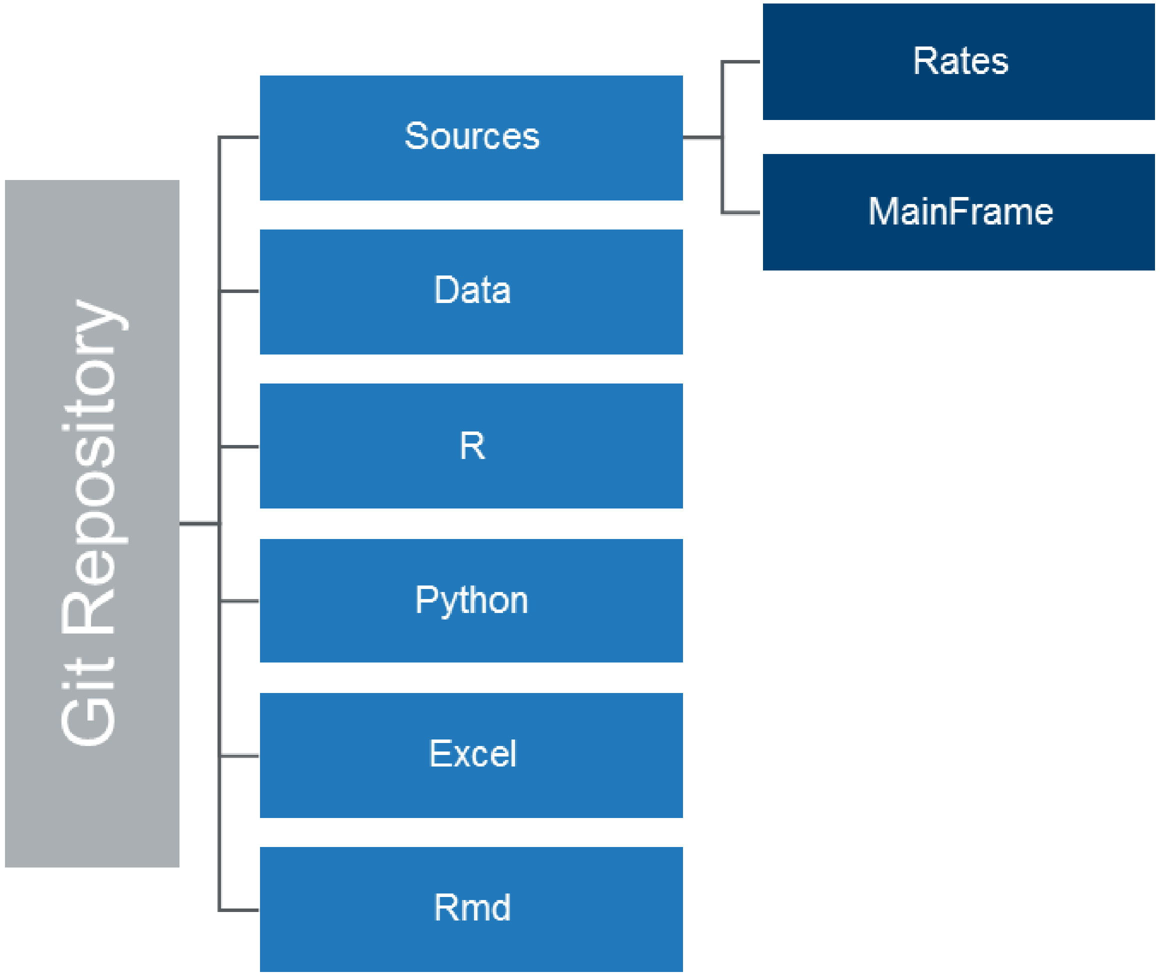
- Three data stores:
  - Legacy data store
  - Current data warehouse
  - Reserving data warehouse
- Reconciliation:
  - Data stores reconciled against each other
  - Data set reconciled against official policy system

Algorithm Selection

Universal Life
Premium Payment
Process



Project structure



Feature engineering and selection

Base Model: P(Pay) = 1/(1+exp(-η)), η = 1.710445. Negative log-likelihood: 1,540,701.

Model 1: P(Pay) = 1/(1+exp(-η)), η = 1.66682267 + 0.00038814 * Funded_Ratio. NLL = 1,539,993.

Coefficients	Estimate	Standard Error	P-Value
(Intercept)	1.66682267	0.00186506	< 2e-16
Funded_Ratio	0.00038814	0.00001054	< 2e-16

Model 2: P(Pay) = 1/(1+exp(-η)), η = 1.3307445 + 0.0977192 * log(1+Funded_Ratio). NLL = 1,535,759.

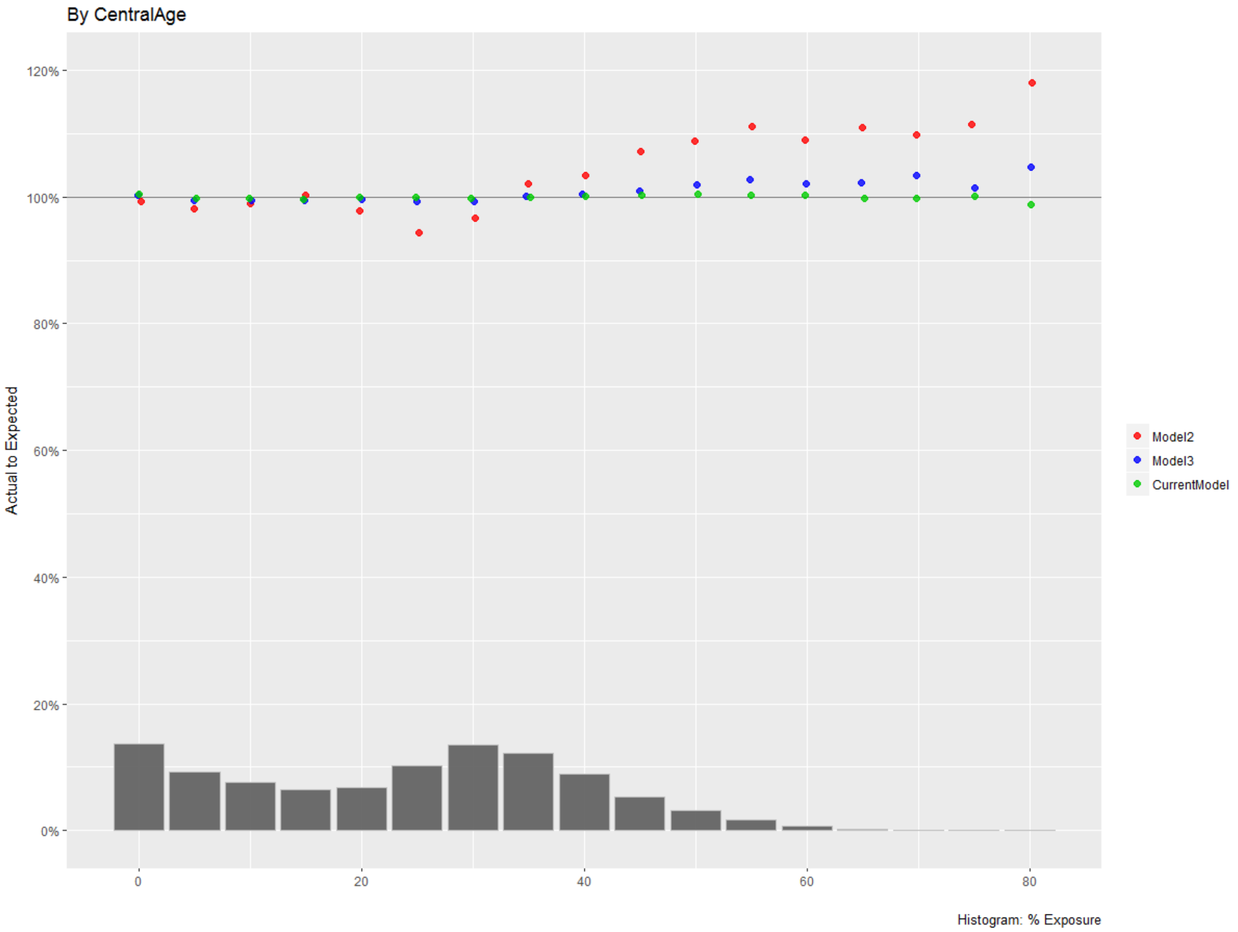
Coefficients	Estimate	Standard Error	P-Value
(Intercept)	1.3307445	0.0039985	< 2e-16
log(1+Funded_Ratio)	0.0977192	0.0009758	< 2e-16

η = 1.331047 + (-0.266503 – 0.993250 I_{NP} + $0.923363 I_{RP}$)* log(1+Funded_Ratio). NLL = 556,227

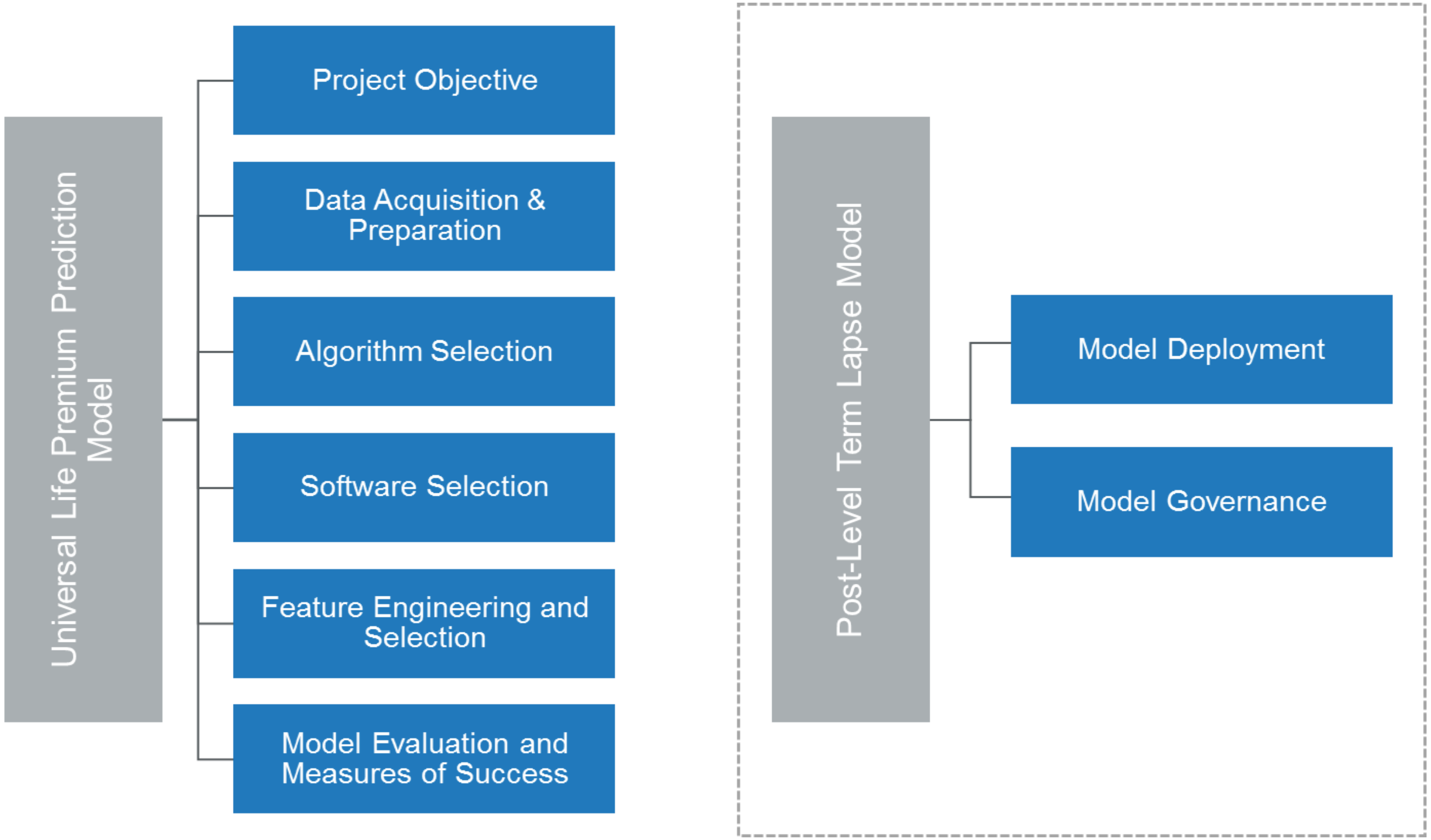
Coefficients	Estimate	Standard Error	P-Value
(Intercept)	1.331047	0.005407	< 2e-16
log(1+Funded_Ratio)	-0.266503	0.001765	< 2e-16
StateNP:log(1 + Funded_Ratio)	-0.993250	0.002858	< 2e-16
StateRP:log(1 + Funded_Ratio)	0.923363	0.001721	< 2e-16

Measures of success

- Actual vs. Expected
- Mean Absolute Percentage Error (MAPE)



Model in production



Model deployment

- Model implemented in company's projection platform
- Company has a dedicated deployment team consisting of programmer actuaries

Model governance

- Model Assumptions Committee
- Executive Finance Committee
- Model Oversight Committee

Model governance

- Model Assumptions Committee
- Executive Finance Committee
- Model Oversight Committee

Questions?