2017 Predictive Analytics Symposium

Session 23, Predictive Modeling Workshops

Moderator:
Rohan Noel Alahakone, ASA, MAAA

Presenter:
Gourab De

DataRobot

How to get started (or move to the next level) of Predictive Analytics

Predictive Modeling Workshops

Gourab De

Head of Practicing Data Science, DataRobot Inc.

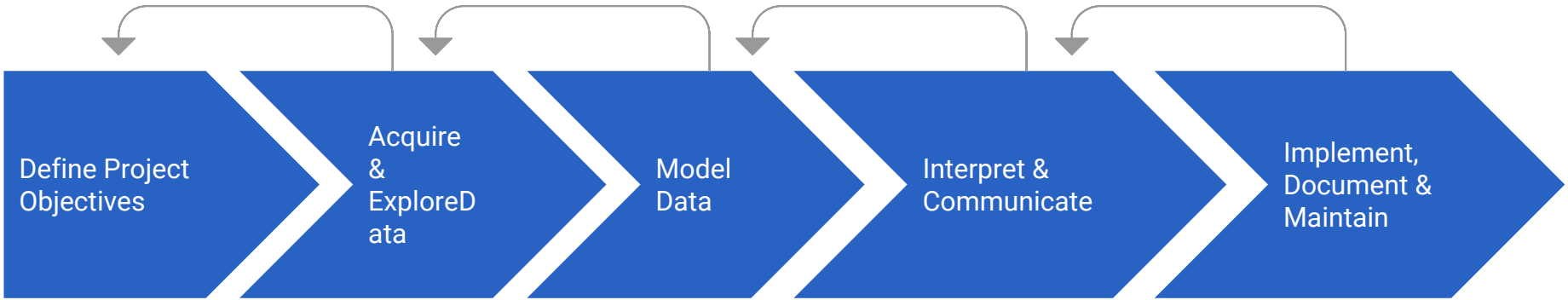
Predictive Analytics Symposium 2017

Society of Actuaries

Chicago, September 14, 2017

SOCIETY OF ACTUARIES

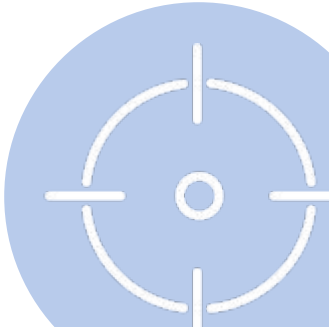
The Machine Learning Life Cycle



Iteration is the key

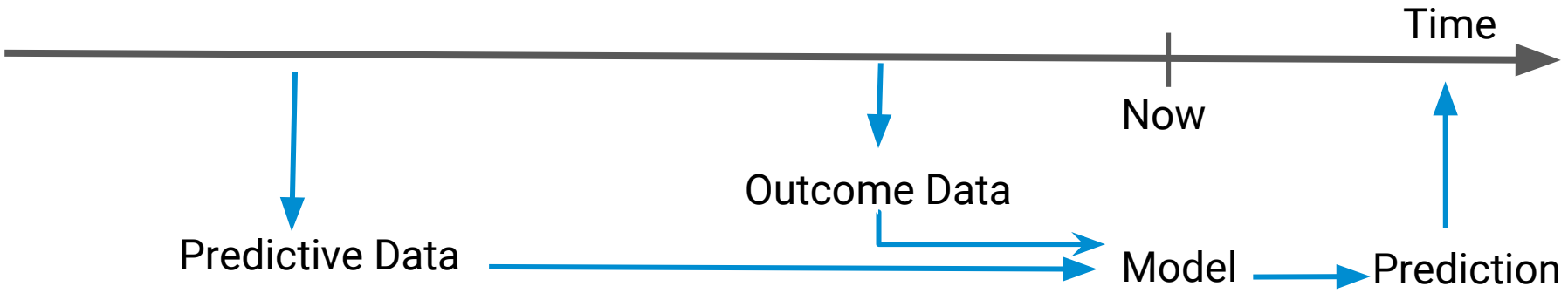
Define Project Objectives

- **Specify business problem**
  - Language of business, include impact statement
- **Define unit of analysis and prediction target**
  - Day or month or year, claimant-level vs. claim-level
  - Survival likelihood in next 5 years, or time to death
- **Prioritize modeling criteria**
  - Accuracy, ease of prediction



Acquire & Explore Data

- **Collect appropriate predictors and sample**
  - Identify time-frame for predictors - past 6 months
  - Ensure timeline is correct



Acquire & Explore Data

- **Collect appropriate predictors and sample**
  - Identify time-frame for predictors - past 6 months
  - Ensure timeline is correct
  - Merge data into single table
- **Conduct exploratory data analysis**

Acquire & Explore Data

- **Collect appropriate predictors and sample**
  - Identify time-frame for predictors - past 6 months
  - Ensure timeline is correct
  - Merge data into single table
- **Conduct exploratory data analysis**
- **Feature engineering**

Model Data

- **Variable selection -**
  - univariate, model-based, manual exclusion
- **Build candidate models**
  - Linear model - strongly dependent on assumption
  - Machine learning models - assumption-agnostic
- **Model validation and selection**
  - Don’t trust a model without comparing out-of-sample - a nice article on cross-validation
  - Compare different approaches on the same test sample

Interpret Model

How well is the model doing overall? - out-of-sample accuracy metrics, ROC curve, lift chart

Are there specific cases where the model is doing poorly? - residual plots

Which predictive variables are most important in determining the outcome? - Feature impact, coefficients

What is the effect of each predictive variable? - Partial dependence

How should we make decisions based on this model? - Reason codes

Implement, Document and Maintain

- [ ] Set up prediction
- [ ] Document modeling process for reproducibility
- [ ] Create model monitoring and maintenance plan

THANK YOU!

Any questions?

You can find me at gourab.de@datarobot.com

https://www.datarobot.com/actuaries/

DataRobot