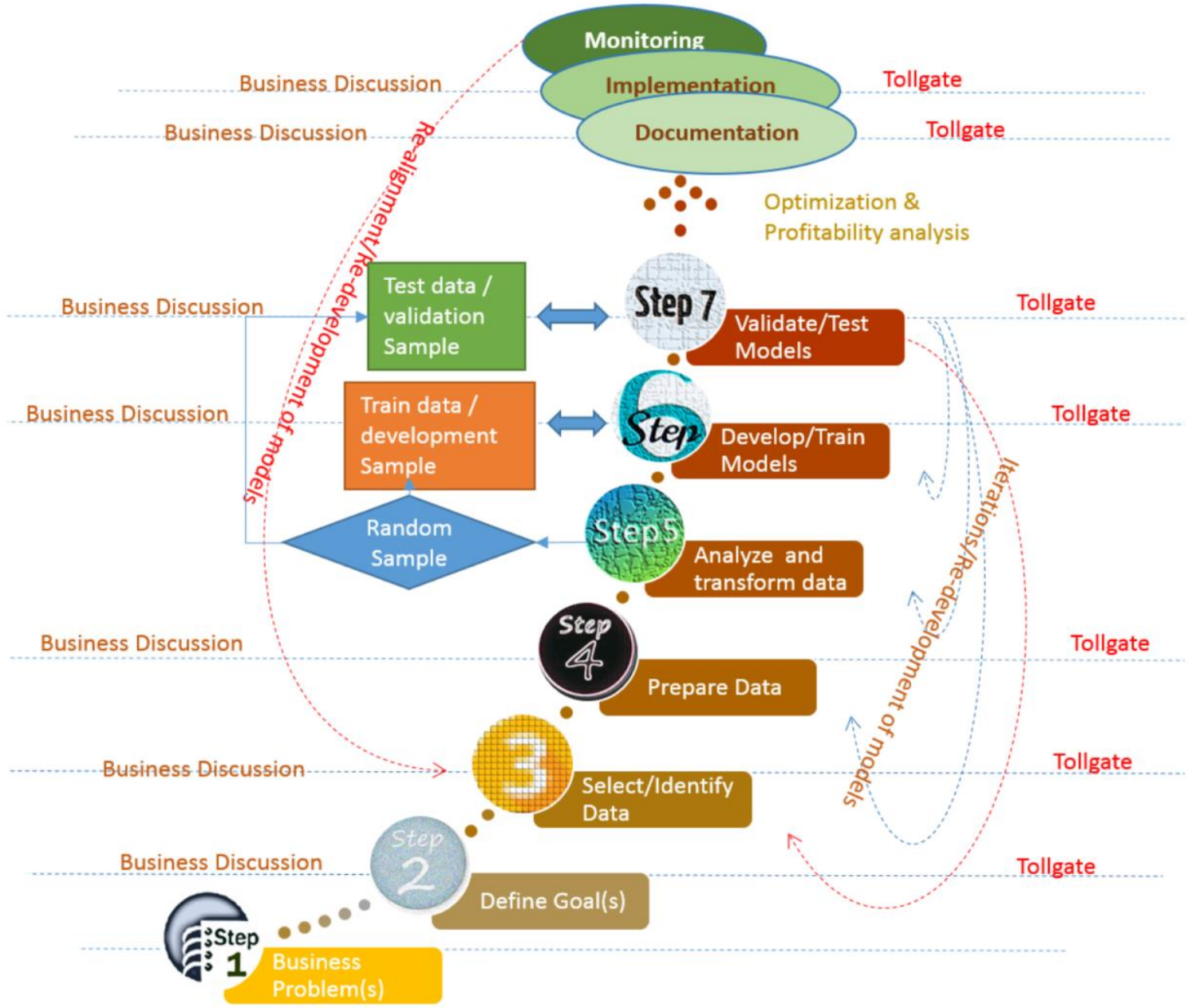
7 Phases Of Predictive Modelling

BY… SAKEEB SHEIKH

7 Phases

- Step 1: Understand Business Objective
- Step 2: Define Modeling Goals
- Step 3: Select/Get Data
- Step 4: Prepare Data
- Step 5: Analyze and Transform Variables. Random Sampling
- Step 6: Model Selection and Develop Models (Training)
- Step 7: Validate Models (Testing), Optimize and Profitability


Monitoring
Implementation
Documentation
Business Discussion
Business Discussion
Tollgate
Tollgate
Optimization &
Profitability analysis
Re-alignment/Re-development of models
Test data /
validation
Sample
Step 7
Validate/Test
Models
Business Discussion
Tollgate
Train data /
development
Sample
Step 6
Develop/Train
Models
Business Discussion
Tollgate
Random
Sample
Step 5
Analyze and
transform data
Iterations/Re-development of models
Step 4
Prepare Data
Business Discussion
Tollgate
3
Select/Identify
Data
Business Discussion
Tollgate
Step 2
Define Goal(s)
Business Discussion
Tollgate
Step 1
Business
Problem(s)

1. Understanding Business Objectives

- Do you want to understand the characteristics of customers?
- Do you want to make unprofitable customers profitable?
- Do you want to understand what driving sales?
- Do you want to win-back lost customers?
- Do you want to increase sales?
- Do you want to reduce customer churn?
- Do you want to reduce cost of production or operation?
- Do you want to target new customers?
- Do you want to identify probable credit default customers?
- Do you want to know X-sell/Upsell opportunities?

Industries

- Retail
- Banking
- Insurance
- Telecom
- Utilities
- Hospitality
- Catalog
- Publishing

2. Define Goals - translate business objective into analytics goal

- Profile Analysis
- Segmentations
- Response Modeling
- Risk Modeling
- Activation
- Cross-Sell and Upsell
- Attrition/Churn Modeling
- Net Present Value(NPV)
- Customer Life Time Value (CLTV)
- etc.

3.Selecting Data for Modeling

Data Type
Demographic
Behavioural
Psychographic

Source of Data	
Internal Source	**External Source**
Customer Data	Survey Data, Research Data, Suppliers, Ratings
Transaction Data	
Other History	Credit Bureau Data, Third Party data, Sellers, Compilers

4. Prepare Data

- Need to prepare data into right format for analysis
- Do initial cleaning up
- Define Variables and Create Data Dictionary
- Joining/Appending multiple datasets
- Validate for correctness
- Produce Basic Summary Reports

5. Analyze and Transform Variables

- Once data is in right shape then perform
  - **univariate analysis**: to check the distribution of each of the variables and features
  - **multivariate analyses**: to check relationships with other variables and with dependent variables

- Based on type of model you may need to transform the variables using one of the approaches
  - Bining approach: create distinct groups
  - Transformation
  - Extreme value (outlier) treatments
  - Missing Value Treatment
  - Dimension Reduction

Random Sampling (Train and Test)

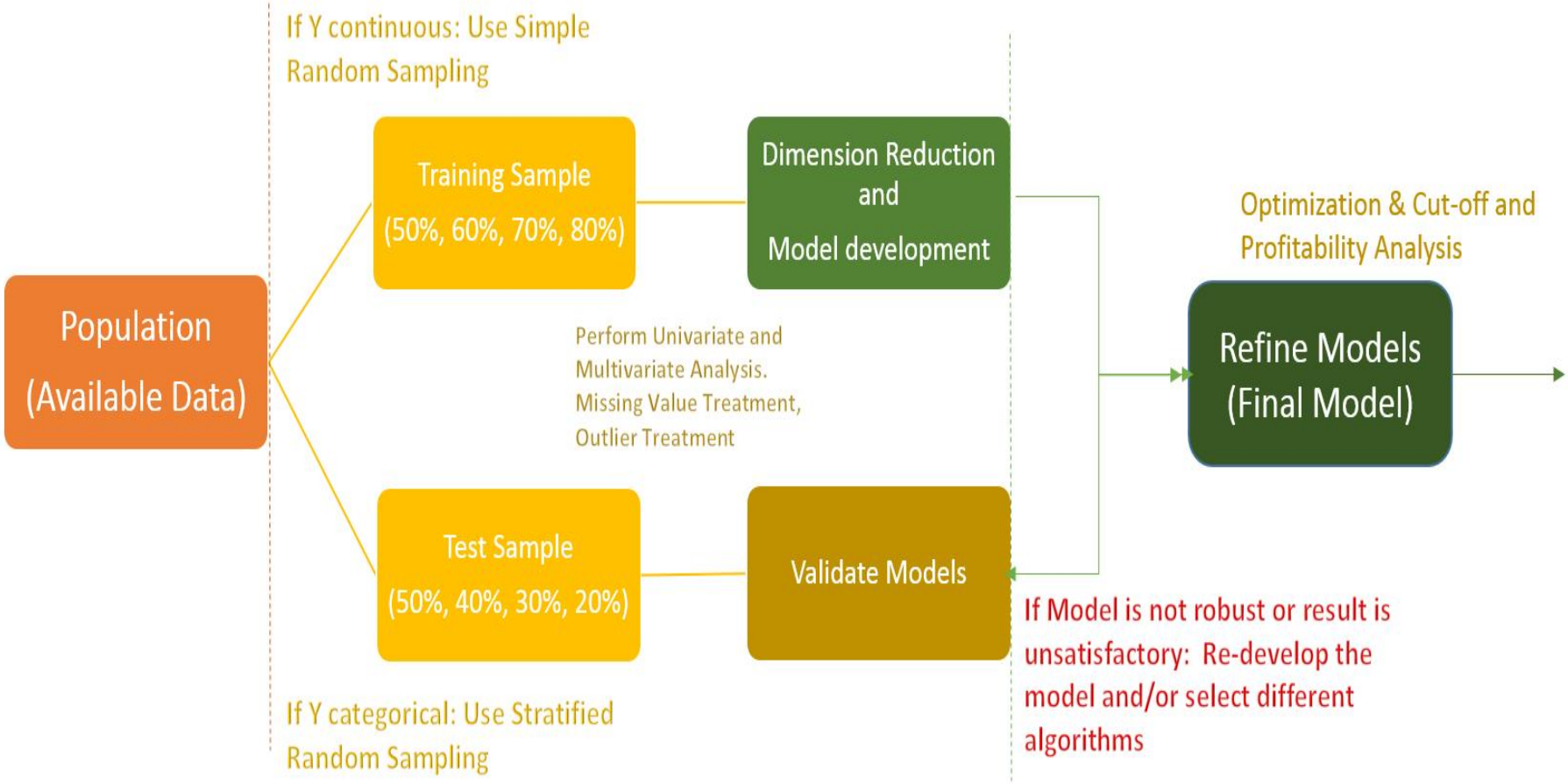
6. Model Selection

- Based on the defined goal(s) (supervised or unsupervised) we have to select one of or combinations of modeling techniques.
  - General linear model
  - Non-Linear Regression
  - Linear Regression
  - Lasso Regression
  - Ridge Regression
  - Non-Negative Garrotte Regression
  - Percentage Regression
  - Quantile Regression
  - Non-parametric regression
  - Logistic Regression
  - Tobit Regression
  - Probit Regression
  - Classification/Decision Trees
  - Random Forest
  - Support Vector Machine (SVM)
  - Distance metric learning

Model Continues…

- Bayesian methods
- Graphical Models
- Neural Networks
- Genetic Algorithm
- The Hazard and Survival Functions
- Time Series Models
- Signal Processing
- Clustering Techniques
- Market Basket Analysis
- Frequent Itemset Mining
- Association Rule Mining etc.

Build/Develop/Train Models



7. Validate/Test Models

- Score and Predict using Test Sample
- Check for the robustness and stability of the model
- Check Model Performance: Accuracy, ROC, AUC, etc.
- Perform Cross Validation to increase accuracy / performance of the models