

Class 5: Model Selection & Prep

Thursday February 22, 2018

Group Paper Timeline

This Week: Finish data acquisition, understanding, and pre-processing

Next Week: Model selection, Feature selection, data preparation

Week of March 5: Model Implementation

Week of March 12: Draft of paper

Spring Break: Buffer time

4 weeks (including today) + Spring Break = 5 Weeks

Today's class: MODEL SELECTION & PREP

Real World Example: Corporate Finance / Credit Risk
-Accounts Receivables / Energy Sector

Assignment #2: Accounts Receivables

Case Study Overview

Corporate Finance / Credit Risk
Accounts Receivables / Energy Sector

Business Understanding

Current & Proposed Specs

	STEPS	Status	Levels	Datasets	Modeling/Scoring Techniques
DATA ACQUISITION & MODELING	Full Credit Scoring Model	Current	Transaction Counterparty (Portfolio?)	Core	Qual/Quant & Regression Based (Scenario Based / Stress Testing?)
	Partial Credit Scoring Model	Under Development	Transaction Counterparty Portfolio	Core Core Proxies Alternative	Qual/Quant Linear (Single/Multi) Beyond Linearity: -Nearest Neighbors -Fine Grained Media Scenario Based / Stress Testing Inc: -Simultaneous correlations -Pre/post recovery Business Processes -Decision Trees
	Hybrid Credit Scoring Model				
	Back-testing (Performance)				
DEPLOYMENT	Technology Integration				CapIQ
	Monitoring				Python Code Based
	Testing & Verification (Tek)				CapIQ

	STEPS	Status	Levels	Datasets	Modeling/Scoring Techniques
DATA ACQUISITION & MODELING	Full Credit Scoring Model	Current	Transaction Counterparty (Portfolio?)	Core	Qual/Quant & Regression Based (Scenario Based / Stress Testing?)
	Partial Credit Scoring Model	Under Development	Transaction Counterparty Portfolio	Core Core Proxies Alternative	Qual/Quant Linear (Single/Multi) Beyond Linearity: -Nearest Neighbors -Business Rules -Graphical Models / Game Theory -Fine Grained Media
	Hybrid Credit Scoring Model				Scenario Based / Stress Testing inc: -Simultaneous correlations -Pre/post recovery
	Back-testing (Performance)				Business Processes -Decision Trees -Business Rules
DEPLOYMENT	Technology Integration				CapIQ
	Monitoring				Python Code Based
	Testing & Verification (Tek)				CapIQ

Credentialization

	Noble	Rating Agencies	Academia	Business/ Partners	Regulators	3 rd Party (\$ Fee / Free)
Structural Concepts						
Modeling Methods inc.						
Quant/Qual Regressions	?	?			?	
Vector Analysis			?		?	
Proximity Analysis			?			
Scenario/Stress Tests Models	?	?				
Data Sources						
Noble Credit Data (3 levels)	?					Local/Global + Automation D&B Rating Agency Actions Country Risk Ratings Bloomberg/Economic Big Data/Google Hose Sentiment Analysis Lexis-Nexis
Ratings & Default Data inc. Ratings conversion Correlations min/max Pre/post recovery losses	?	?				
Scenario & Stress Tests Data	?	?				
Proximity Data	?		?	?	?	

Data Acquisition: Current & Needs

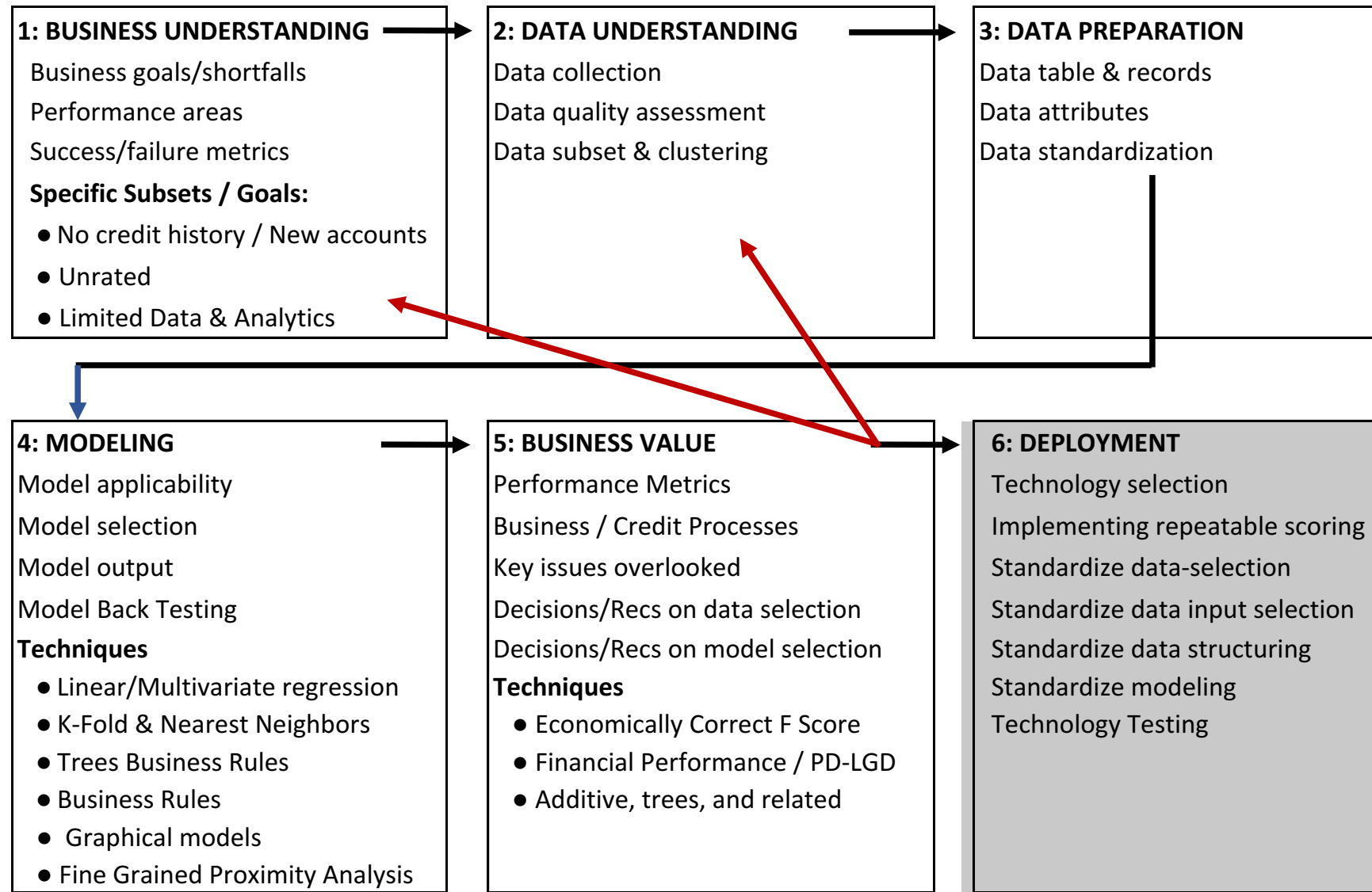
	Noble	Rating Agencies	Academia	Business/ Partners	Regulators
Noble Credit Data (3 levels)	?				
Ratings & Default Data	?	?			
Scenario & Stress Tests Data	?	?			
Proximity Data	?		?	?	?

3 rd Party	Available	Integrated	Real Time	Data Type
CapIQ	Yes	Yes	?	
D&B	Yes	Yes	No	
Rating Agency Actions	Yes	Yes	No	
Country Risk Ratings	?	?	N/A	
Bloomberg/Economic	?	?	N/A	
Big Data/Google Hose	No	No	N/A	
Sentiment Analysis	No	No	N/A	
Lexis-Nexis	?	?	N/A	

Data Classification

See Excel File_CM Tables Session 2

Data & Modeling Process



Modeling

4: MODELING

Model applicability

Model selection

Model output

Model Back Testing

Techniques

- Linear/Multi/Poly
- K-Fold & Nearest Neighbors
- Trees Business Rules
- Graphical models
- Fine Grained Proximity Analysis

Hastie ISL

F score 2.1 / p.17-21

Linear: 3.1-3.3 / p.61-75

Libraries 3.6.1-3 / p.100-109

Refresher R: 2.3/42

Polynomial: 7.1-3 / p. 266-270

Classification:

KNN (K-Nearest Neighbors): 4.6.5 / p.163-164

Examples: 4.7 / 168-174 (+Stock Market)

Resampling:

K-Fold Cross Validation: 5.3.3. / p.193

Examples: 5.1 / p.197-202

Regularization:

Lasso 6.2.2 / p.219

Examples: 6.8 / 259

Trees

Regression/Classification: 8.1.1-8.1.4 / p.303-313

Examples: 8.3.1-8.3.2, p. 323-327

Business Value



5: BUSINESS VALUE

Performance Metrics

Business / Credit Processes

Key issues overlooked

Decisions/Recs on data selection

Decisions/Recs on model selection

Techniques

- Economically Correct F Score
- Financial Performance / PD-LGD
- Additive, trees, and related