

Teaching Predictive Modeling with JMP Pro 13

Mia Stephens – mia.stephens@jmp.com



THE
POWER
TO KNOW®

What is Analytics?

“...an encompassing and multidimensional field that uses mathematics, statistics, predictive modeling and machine-learning techniques to find meaningful patterns and knowledge in recorded data.”

- Descriptive statistics
- Predictive analytics
- Prescriptive analytics

From SAS Analytics:

http://www.sas.com/en_us/insights/analytics/what-is-analytics.html

Flavors of Analytics

- **Descriptive statistics**

- Understand what happened and why
- Exploratory and explanatory analysis and modeling
- *Classical statistics, inference, and modeling*

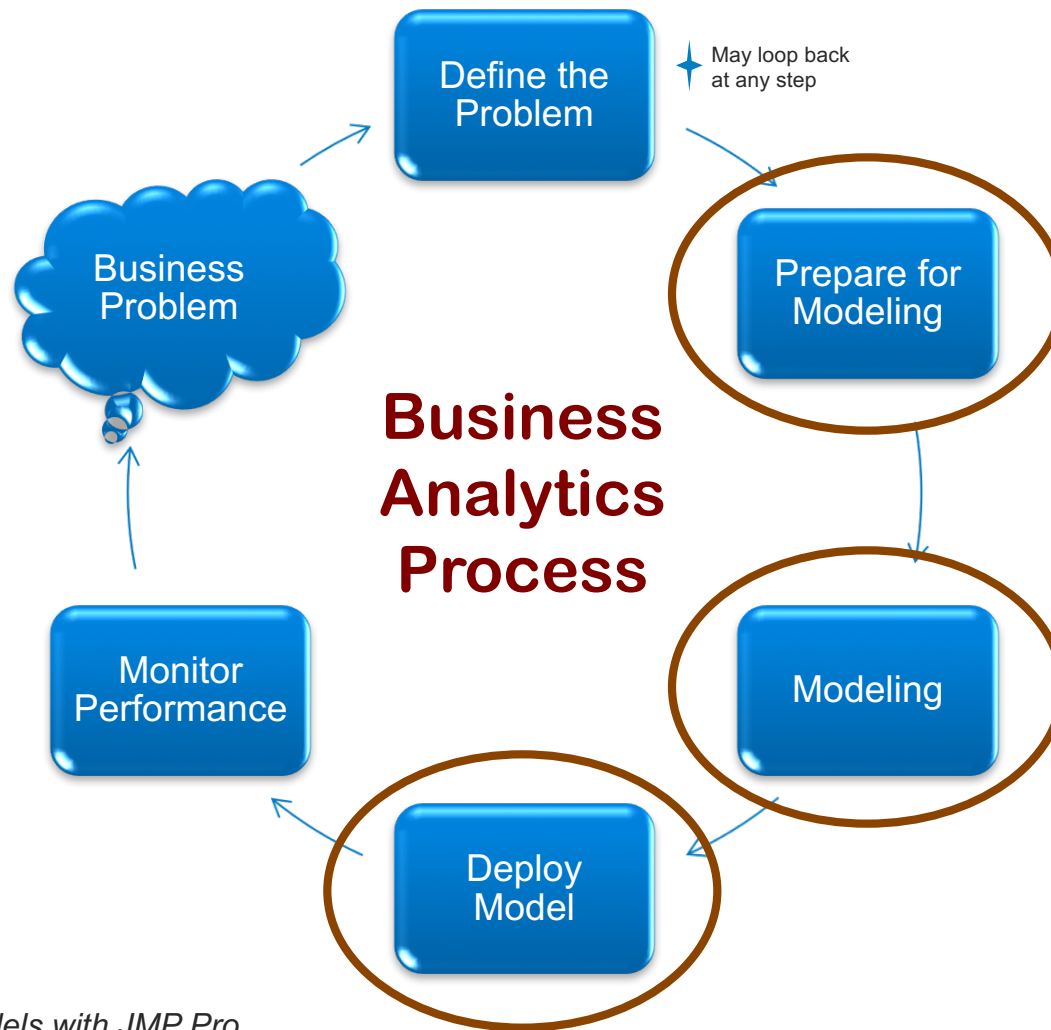
- **Predictive analytics**

- Using past data and predictive algorithms to determine what will happen next
- *Predictive modeling*

- **Prescriptive analytics**

- Answering the question of what to do, by providing information on optimal decisions based on the predicted future scenarios
- *Simulation and optimization*

The Business Analytics Process



From *Building Better Models with JMP Pro*,
Grayson, Gardner and Stephens, 2015.

Data Preparation: Activities and Tools

Key Activities:

- Determine which data are needed
- Compile (or collect new) data
- Explore, examine and understand data
- Assess data quality
- Clean and transform data
- Define features
- Reduce dimensionality
- Create training, validation and test sets

Key Tools:

- SQL/Query
- Data table structuring - join, concatenate, update, stack, summarize,...
- Summary statistics and graphical displays, interactive tools and filtering Multivariate procedures (clustering, PCA,...)
- Transformations, creating derived variables
- Missing data utilities, outlier analysis, recoding, binning
- Creating holdout set(s)

Modeling: Activities and Tools

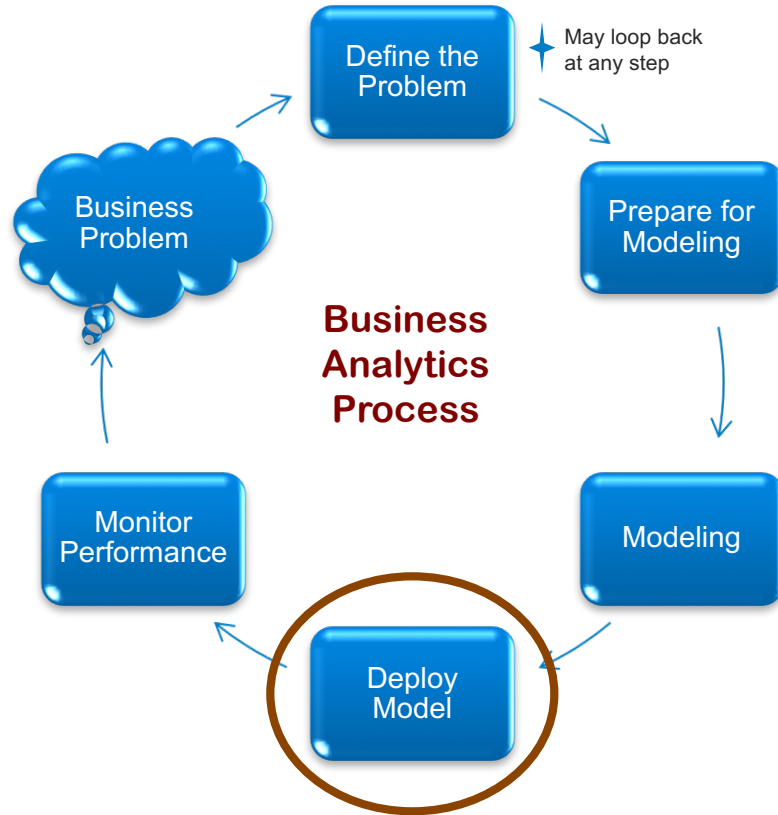
Key Activities:

- Choose the appropriate modeling method or methods
- Fit one or more models
- Evaluate the performance of each model using validation statistics (misclassification, RMSE, RASE, Rsquare)
- Choose the best model or set of models to address the analytics problem (and ultimately the business problem)
- **Create ensemble models

Key Tools:

- Multiple Regression
- Logistic Regression
- Naïve Bayes
- kNN
- Classification and Regression Trees
- Bootstrap Forests and Boosted Trees
- Neural Networks
- Generalized Linear Models
- Survival Models
- Forecasting/Time Series
- Model Comparison
- Text Mining

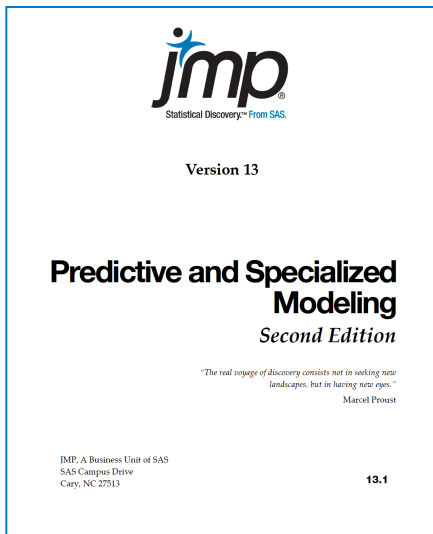
Deploy Models: Activities and Tools



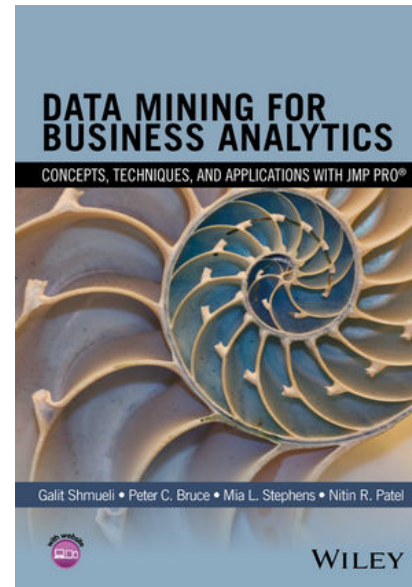
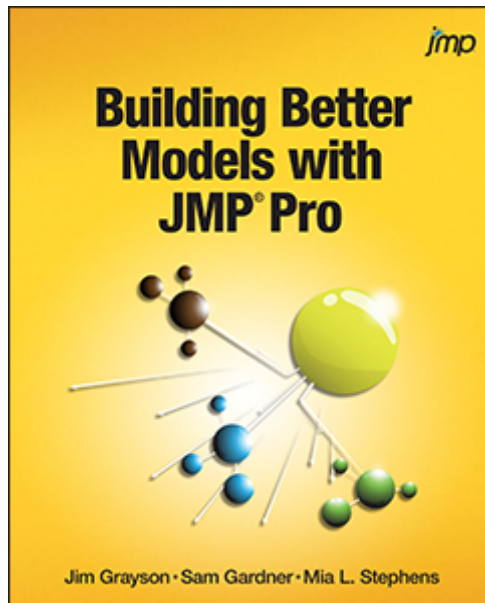
Key Activities and Tools:

- Communicate results (graphs and profilers, summaries, explore “what if” scenarios)
- Deliver the model and model results to the business or internal customers (Formula Depot, write Python, SAS, Java Script, SQL, C)
- Assist in applying model insights and implementing the model
- Document the project
- Follow up with the business sponsor to close out the project

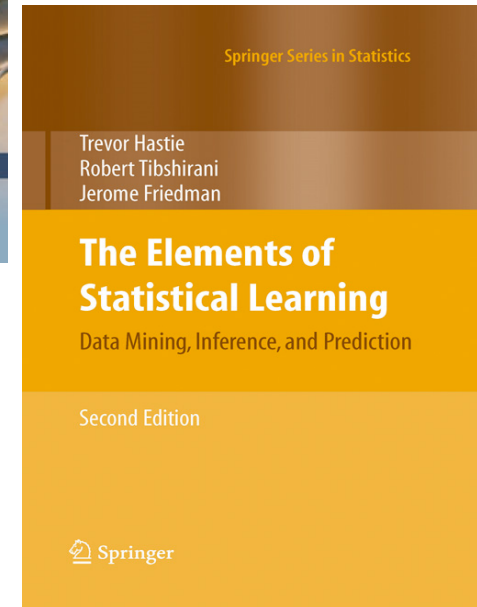
For more information...



Help > Books



jmp.com/getbooks



jmp.com/getbooks

ESL: <http://statweb.stanford.edu/~tibs/ElemStatLearn/>

For more information...

- jmp.com/academic
 - Learning library (guides and videos)
 - Live and recorded webinars
 - Case studies
 - Books for teaching with JMP
 - Tools for teaching statistical concepts
- jmp.com/courses – Predictive Modeling
- Community.jmp.com
 - Discussion forum
 - File exchange – examples and sample datasets
 - Data mining tools add-ins
 - More teaching and learning resources



Discussion and Q&A

Mia.stephens@jmp.com
jmp.com/academic