

Data Science: The New Hybrid Discipline

Tom Fomby

Department of Economics

SMU

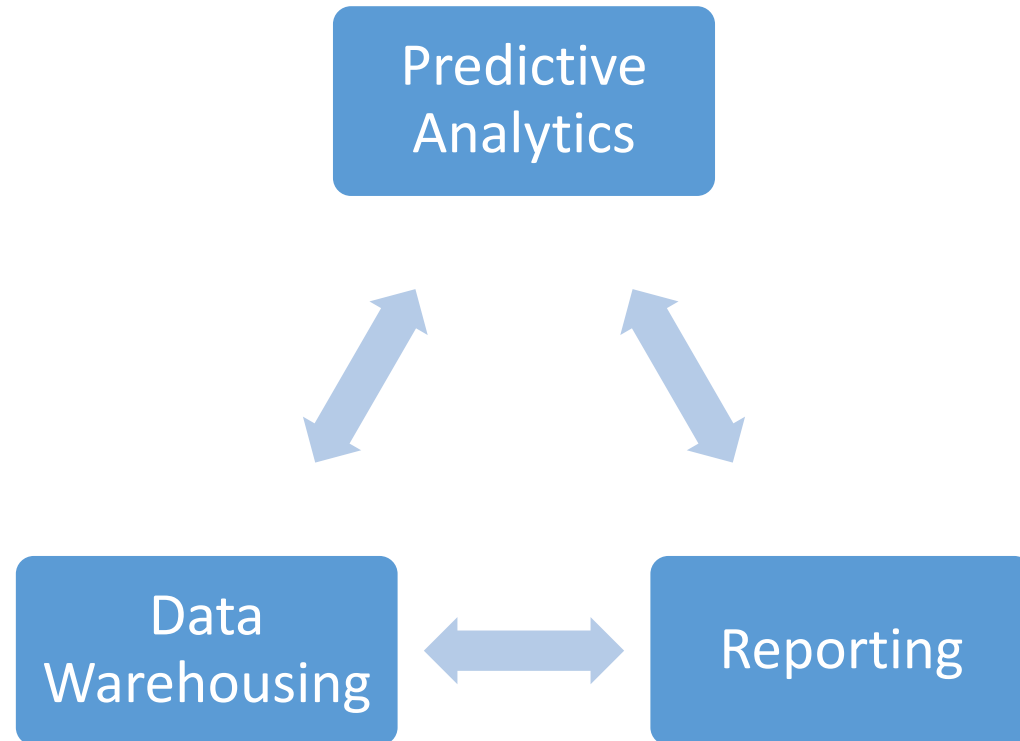
DABE Presentation

Friday, February 19, 2016

DATA SCIENCE: What is it?

- It is an **interdisciplinary field** involving training in computer programming, statistical methods, machine learning tools, and visual presentation with good communications skills. Data Scientists are important providers of insights to decisionmakers in business, government, medicine and many other fields that result in more efficient and effective decisions.
- **Computer and Statistical skills**
- **Machine learning tools**
- **Graphical Software and Presentation Skills**
- **A Knowledge of the Economic Way of Thinking is a very important attribute of a Data Scientist!**

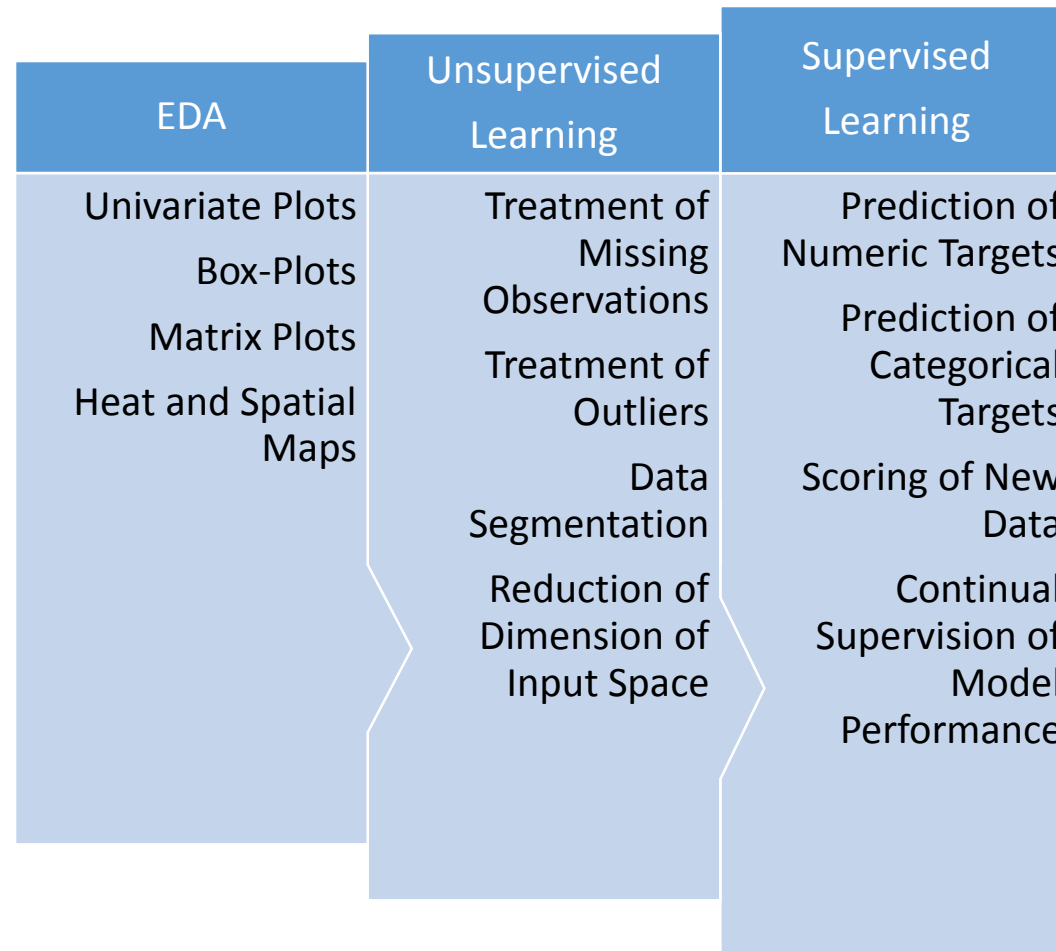
Successful Implementation of Predictive Analytics into company operations requires a combination of three Basic Core Competencies



Some Specifics on Skills

- Multiple Linear Regression (OLS, WLS, Time Series and Spatial Regressions)
- Generalized Linear Modeling (Probit, Logit, Multinomial Logit/Probit, Count, Cox Proportional Hazard models, Truncated and Censored models)
- Time Series Modeling Expertise (Seasonal Adjustment, Box-Jenkins, Exponential Smoothing, Vector Autoregressions)
- Applied Multivariate Statistical Analysis (Clustering, Principal Components, Discriminant Analysis)
- Training in Machine Learning Tools (CART, CHAID, SVM, ANN, K-Nearest-Neighbors, LASSO, Association Rules)
- Computer Usage and Programming Skills (SPSS Modeler, SAS Enterprise Miner, Matlab, Mathematica, R, STATA, SAS, EVIEWS)

Core Tasks of Predictive Analytics



A Central Tenet of Predictive Analytics:

Prediction is More Important than
Hypothesis Testing

A Flow Diagram of a Typical
University Course
in Data Mining

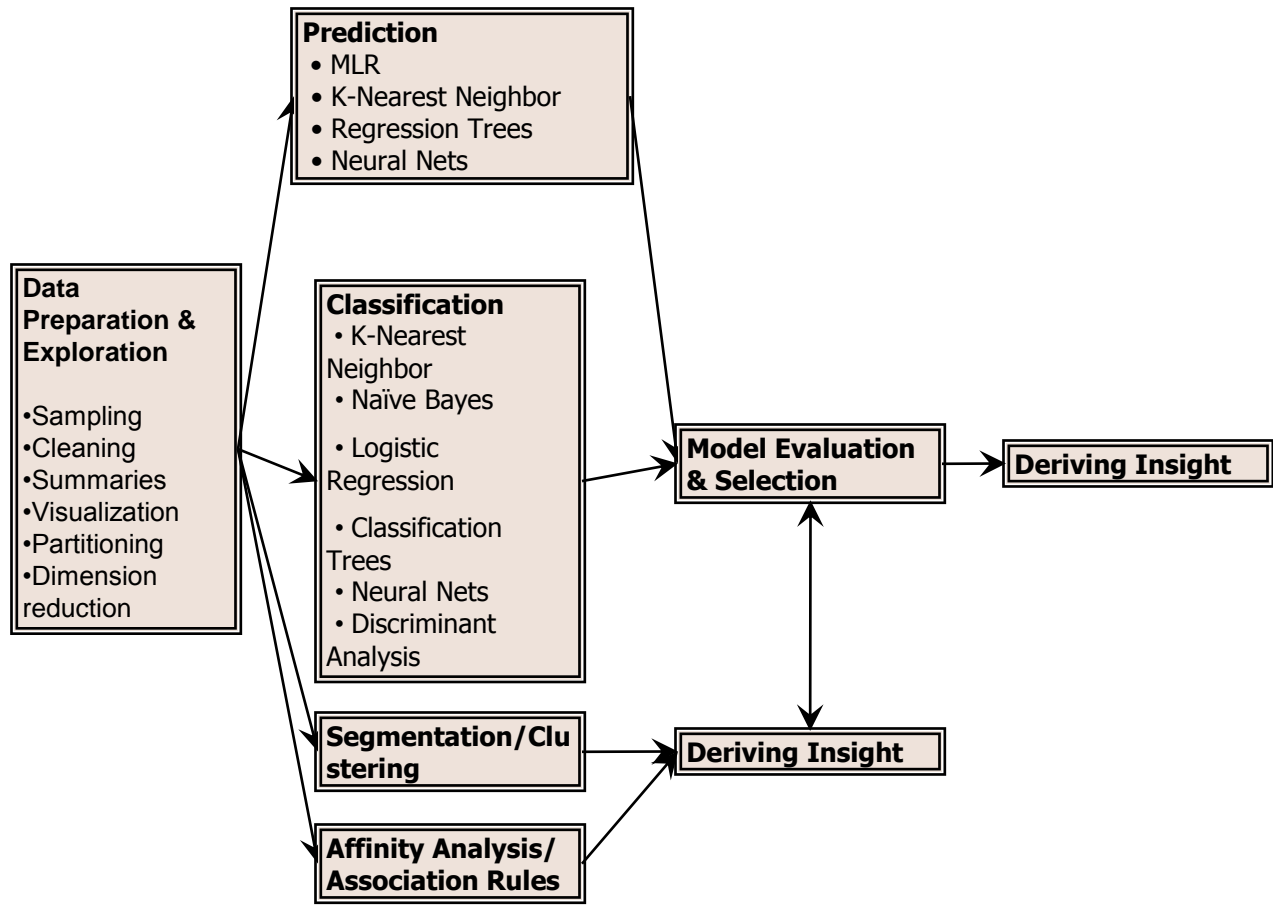
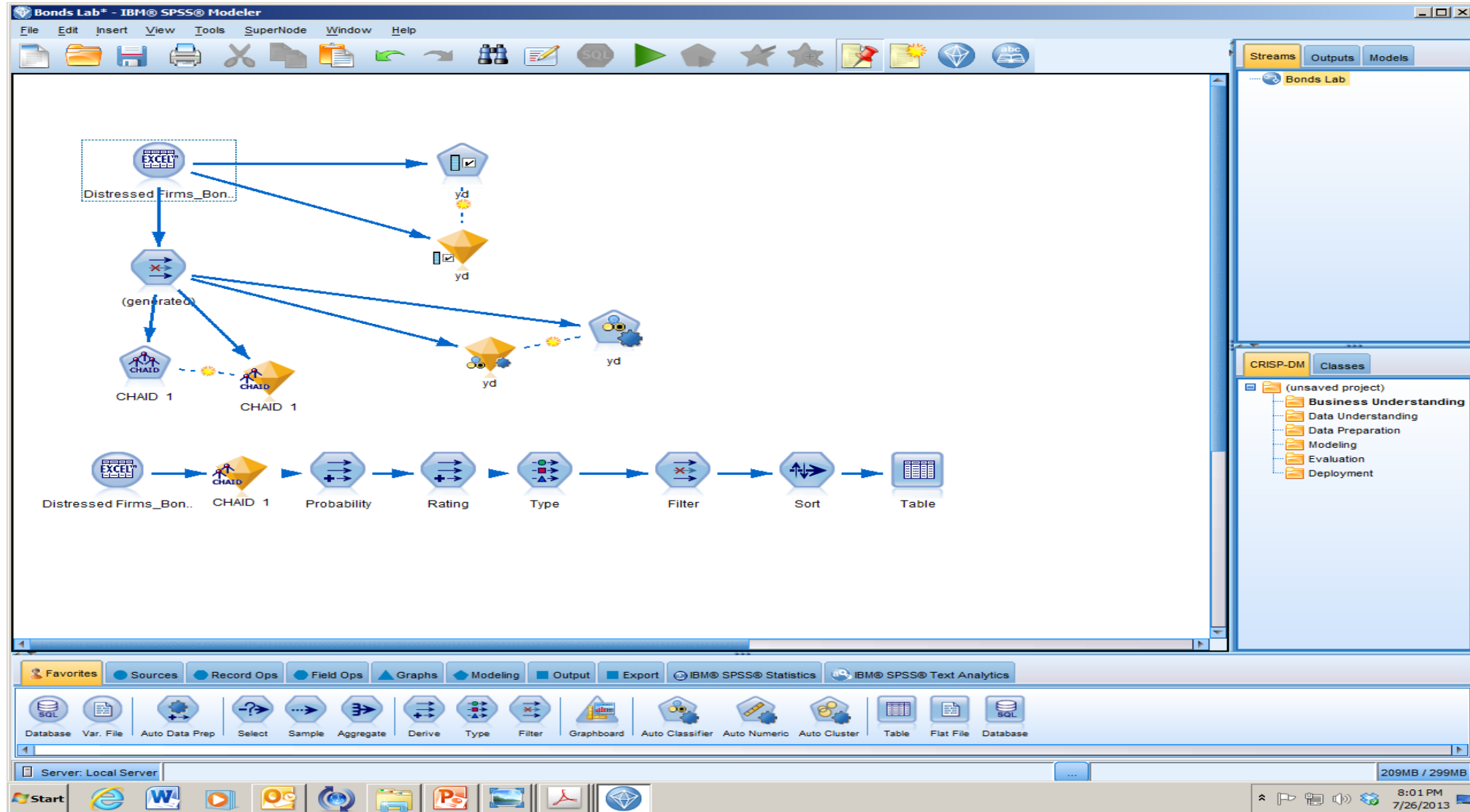


Figure 1.2: Data mining from a process perspective

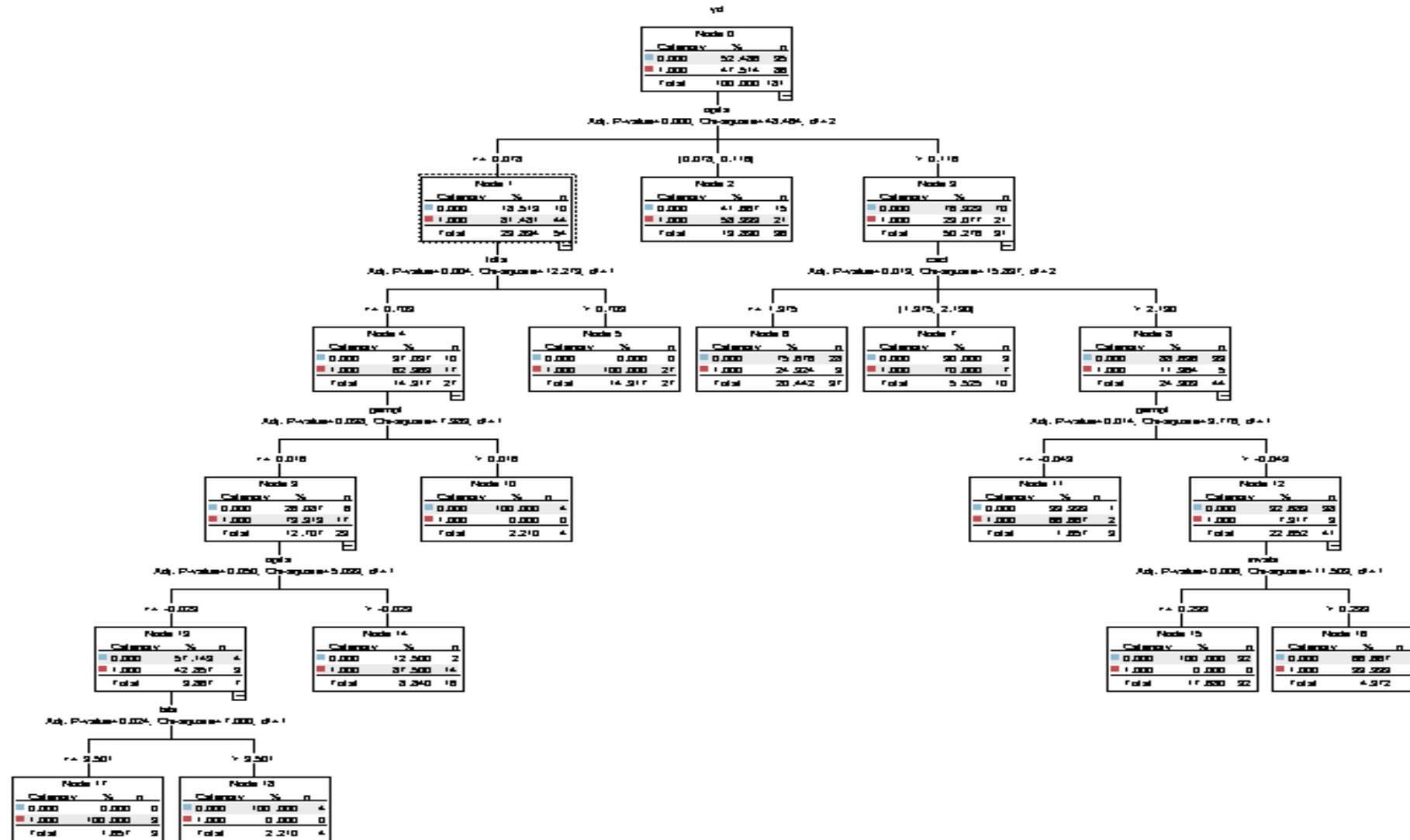
G. Samueli, N. R. Patel and P.C. Bruce. *Data Mining for Business Intelligence (2007)*.

SOME ODD LOOKING THINGS TO ECONOMISTS

A Typical SPSS Modeler Stream

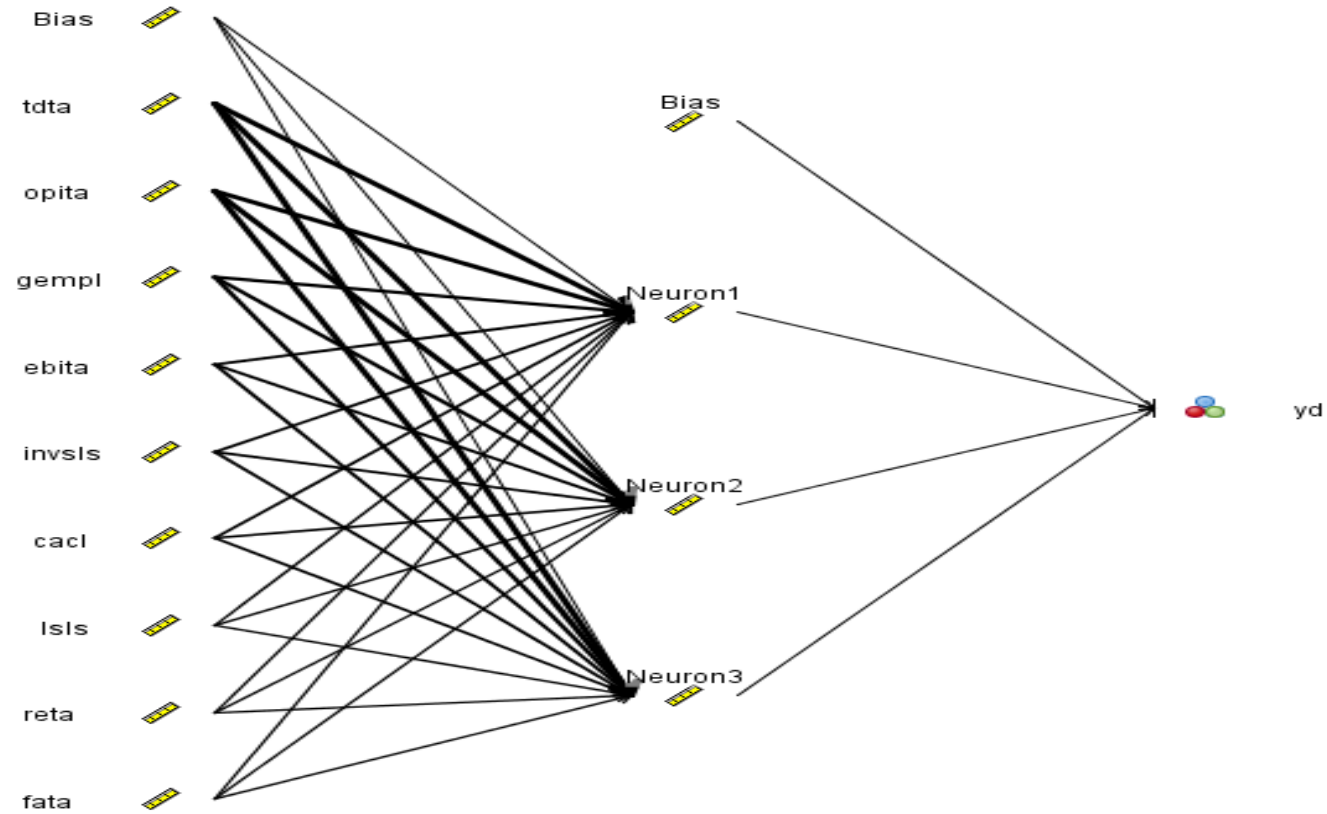


A Decision Tree – A Contribution of Herbert Simon, Nobel Prize in Economics, 1978



An Artificial Neural Network

Network



Least Important

Most Important

Some Important Applications of Predictive Analytics

- Individualized Medicine
- Customer Segmentation
- Customer Retention
- Target Marketing
- Credit Scoring
- Default Analysis
- Crime Analytics
- Government Planning
- Website Design
- Sentiment Analysis
- Text Analytics
- Fraud Detection
- Warranty Analysis

Data Science Programs are Proliferating

- SMU has four Data Science programs, one of them being online
- Other On-line programs: Northwestern University, Berkeley, NYU to name a few
- The University of Rochester has recently invested \$100 million in setting up a Data Science Institute
- Temple University has just established a Big Data Institute working with local companies and research institutions

The Curriculum of Economics is Evolving: The Statistization of Economics as compared to the Mathematization of Economics

- More Economics Majors than ever before and growing rapidly
- More Applied Courses and Heavier Computer Usage
- More Applied MA/MS Economics Programs
- A Multitude of Computer Packages
- Versatility of Economists is a hallmark
- The Internet: Economists are practicing more experimental design
- More Internships
- More supplementary E-learning courses in computer science and predictive analytics
- More short courses in SAS, etc. (encouraging students to become certified in SAS for resume building purposes)
- NABE is now offering the CBE certificate validating the curriculum of Applied Masters programs in Economics as compared to the MBA

The Future Looks Bright for Economists
both in Academic and Non-academic
Positions!

THANK YOU!

ANY QUESTIONS?