



Business Analytics

IBM
Software
Solutions
Group

Predictive Analytics Workshop using IBM SPSS Modeler

Christine Lopez – Technical Sales Specialist

November 19, 2014



Objectives

Smarter Software for a Smarter Campus



Agenda

Welcome and Introductions

The Predictive Analytics Process

Exercise: Navigating IBM SPSS Modeler

Exercise: Predictive in 20 Minutes

Break

Data Mining Methodology and Application

Exercise: Data Mining Techniques and Deployment

Wrap-up

Purpose of the Workshop

- Introduction to predictive analytics and data mining
- Stimulate thinking about how data mining would benefit your institution
- Demonstrate ease of use of powerful technology
- Get experience in “doing” data mining
- See examples of how other organizations are benefitting from deploying predictive analytics

The Smarter Campus



What if you knew which **prospective students** to recruit?

What if you could **detect financial aid fraud**?

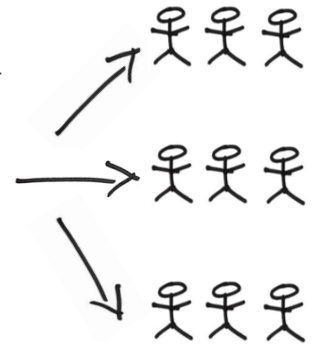
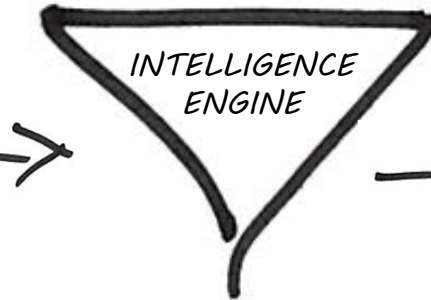
What if you could identify which **students** were at - risk of dropping out?

What if you knew which **alumni** would **donate**...and when?

What if you could detect **campus crime** before it happened?

CRM DATA

- + NAME
- + CONTACT INFO
- + PAST INTERACTIONS
- + TRANSACTIONS



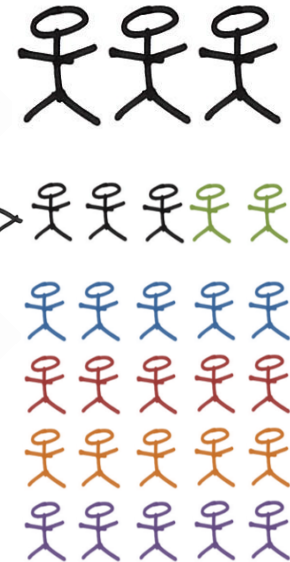
PUBLIC RECORDS
LMS SURVEYS



???

CRM DATA

- + NAME
- + CONTACT INFO
- + PAST INTERACTIONS
- + TRANSACTIONS



IBM SPSS Modeler
A Quick Overview



IBM SPSS Modeler at a Glance

- **Comprehensive predictive analytics platform**
- **All types of users at multiple analytical abilities**
- **Integrated deployment, optimization and decision management capabilities**
- **A visual interface with built-in guidance**
- **Structured and unstructured data**
- **Deployed on a desktop or integrated within operational systems**
- **Bring predictive intelligence to a single user or an entire enterprise**

Predictive in 20 Minutes

A Quick Exercise

3

Exercise: Predictive in 20 Minutes

Goal:

- Create a model to identify students likely to enroll

Approach:

- Use student data which contains demographic and academic information
- Define which fields to use
- Choose the modeling technique
- Automatically generate a model to identify who are likely to enroll
- Review results

Why?

- For enrollment management implications, by proactively identifying and targeting students likely to enroll

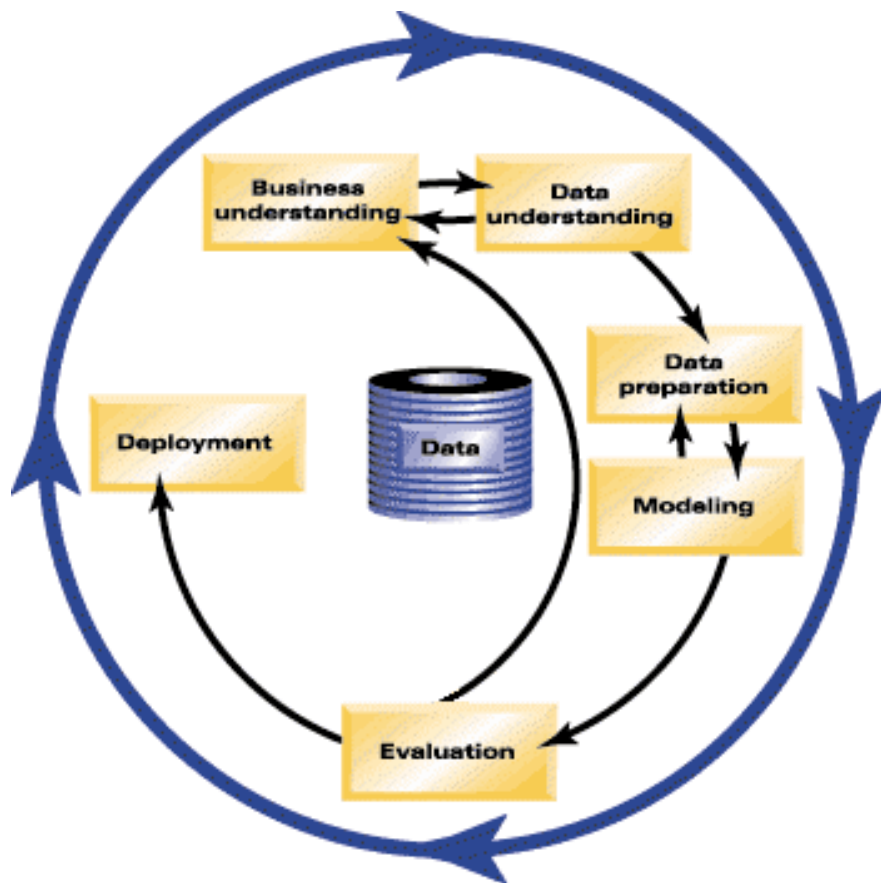
Break - Please Return in 15 Minutes



IBM SPSS Modeler
One Analytical Workbench – Endless Techniques



Data Mining Methodology



- **Cross-Industry Standard Process Model for Data Mining**
- **Describes Components of Complete Data Mining Project Cycle**
- **Shows Iterative Nature of Data Mining**
- **Vendor and Industry Neutral**

Data Mining Techniques

Technique	Usage	Algorithms
Classification (or prediction)	<ul style="list-style-type: none">• Used to predict group membership (e.g., will this employee leave?) or a number (e.g., how many widgets will I sell?)	<ul style="list-style-type: none">• Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc.

Data Mining Techniques

Technique	Usage	Algorithms
Classification (or prediction)	<ul style="list-style-type: none">• Used to predict group membership (e.g., will this employee leave?) or a number (e.g., how many widgets will I sell?)	<ul style="list-style-type: none">• Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc.
Segmentation	<ul style="list-style-type: none">• Used to classify data points into groups that are internally homogenous and externally heterogeneous• Identify cases that are unusual	<ul style="list-style-type: none">• Auto Clustering, K-means, etc.• Anomoly detection

Data Mining Techniques

Technique	Usage	Algorithms
Classification (or prediction)	<ul style="list-style-type: none">• Used to predict group membership (e.g., will this employee leave?) or a number (e.g., how many widgets will I sell?)	<ul style="list-style-type: none">• Auto Classifiers, Decision Trees, Logistic, SVM, Time Series, etc.
Segmentation	<ul style="list-style-type: none">• Used to classify data points into groups that are internally homogenous and externally heterogeneous.• Identify cases that are unusual	<ul style="list-style-type: none">• Auto Clustering, K-means, etc.• Anomaly detection
Association	<ul style="list-style-type: none">• Used to find events that occur together or in a sequence (e.g., market basket)	<ul style="list-style-type: none">• APRIORI, Carma, Sequence

Additional Data Mining Techniques

Technique	Usage	Algorithms
Text Analytics	<ul style="list-style-type: none">• Used to discover patterns resident in text or other unstructured data (e.g., sentiment analysis)	<ul style="list-style-type: none">• Natural Language Processing• Parts of Speech Analysis
Entity Analytics	<ul style="list-style-type: none">• Used to determine which cases are likely the same actor, and which seemingly identical cases are actually independent	<ul style="list-style-type: none">• Context Accumulation
Social Network Analysis	<ul style="list-style-type: none">• Used to uncover associations which may exist between cases, and identify central or influential actors	

IBM SPSS Modeler
Segmentation Modeling

5

Segmentation Modeling

Goal:

- Discover natural groupings or clusters of alumni donors

Approach:

- Alumni data from a university
- Define which fields to use
- Use K-Means Clustering to generate a model to group alumni
- Appendix: Use these clusters to predict donation

Why?

- Better alumni understanding (demographics, socio-economic etc)
- Tailored messages for each group/segment
- Personal and more relevant for alumni
- Institutional Planning

IBM SPSS Modeler
Classification Modeling

6

Classification model

Goal:

- Identify students likely to persist

Approach:

- Use student performance scores and other demographics
- Define which fields to use
- Use the Auto Classifier to choose the appropriate modeling technique
- Review results

Why?

- Identify students likely to persist into their second year
- Conversely, same methods can be used to identify students at risk of attrition (or prisoners at risk of recidivism, or patients likely to respond to treatment)

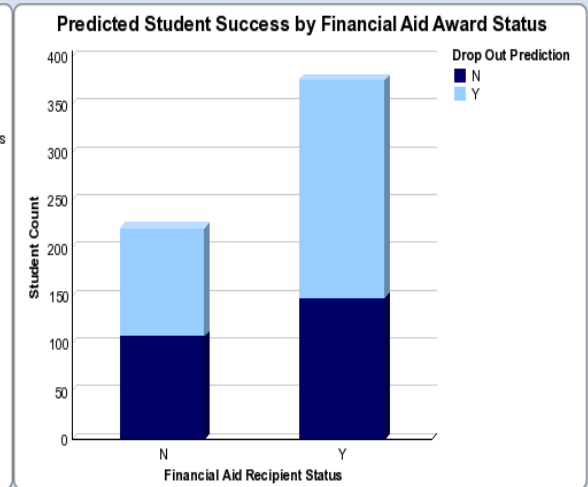
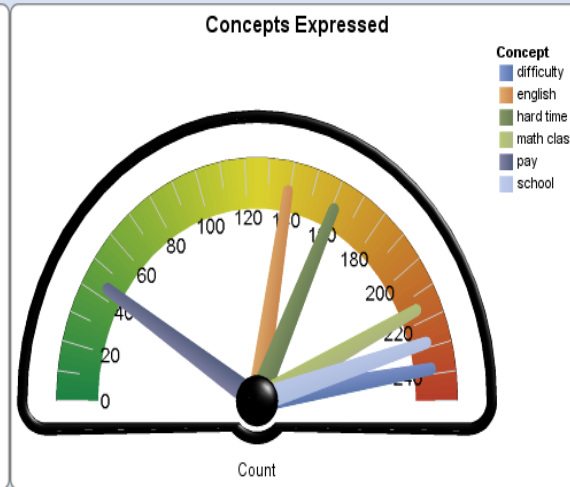
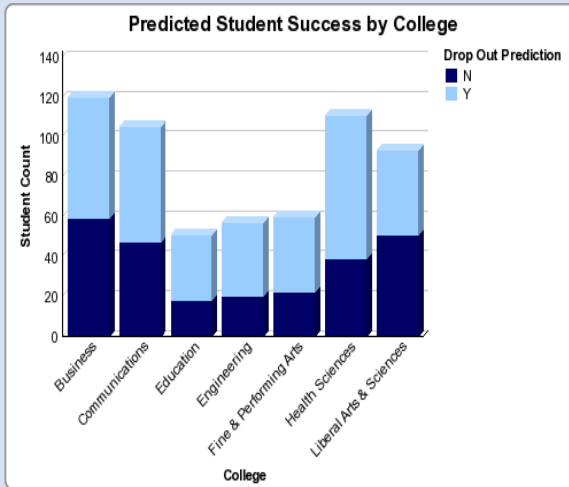
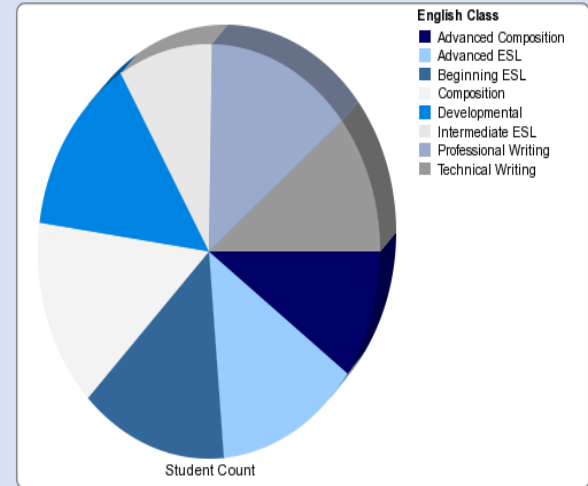
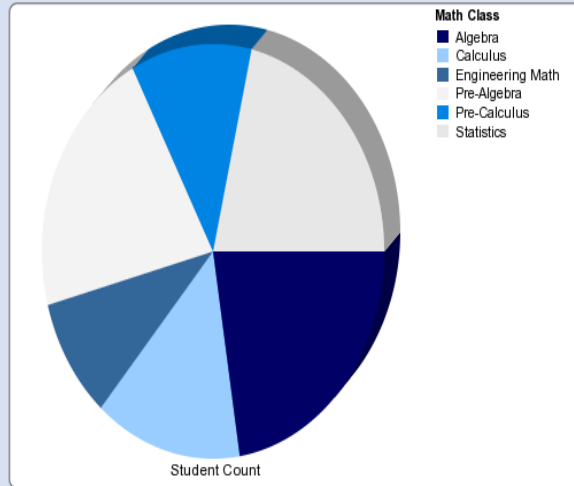
Deployment
Many Options



Enrollment Counselor View - Somewhat Likely to Enroll Students

Student ID	First Name	Last Name	Ethnicity	Enroll Prediction	Propensity to Enroll	Alumni Meeting	Facebook Page Member	Overnight Campus Visit	Financial Aid Awarded	Mail Packet
600361267	GILBERT	MCCORMICK	Hispanic	Y	52.29%	N	Y	Y	26% - 50%	N
593017526	JOSH	JENKINS	Unknown	Y	65.79%	N	Y	Y	1% - 25%	Y
554751518	KATHRYN	MORRISON	White	Y	65.79%	N	Y	Y	1% - 25%	N
314905183	RANDY	LEWIS	White	Y	65.79%	N	Y	Y	1% - 25%	N
301841846	LAURA	ROCHA	White	Y	65.79%	N	Y	Y	1% - 25%	Y
369908296	RYAN	MCCORMICK	White	Y	66.18%	Y	Y	N	0%	Y
597308314	CARLOS	SMITH	Hispanic	Y	67.24%	Y	Y	N	0%	Y
293980526	RAYVIN	MURPHY	Black	Y	67.24%	Y	Y	N	0%	Y
275824987	JASMINE	MURPHY	Black	Y	67.24%	Y	Y	N	0%	Y
260872895	MOHAMED	ROBERTSON	Black	Y	67.24%	Y	Y	N	0%	Y
252856915	VICTOR	MITCHELL	Black	Y	67.24%	Y	Y	N	0%	N
158802483	KATRINA	JOHNSON	Black	Y	67.24%	Y	Y	N	0%	Y
118643872	DEREK	JENKINS	Black	Y	67.24%	Y	Y	N	0%	N
322705060	FERNANDO	MONROE	Unknown	Y	68.60%	N	Y	Y	26% - 50%	N
209600042	JENNIFER	MALONEY	White	Y	75.00%	N	Y	Y	1% - 25%	N

President View



Why IBM SPSS?

8

Workshop takeaways

Easy to use, visual interface

- Short timeframe to be productive with actionable results
- Does not require knowledge of programming language
- No proprietary data formats
- Open architecture

Business results focused

- Leverages the investments already made in technology
- Cost effective solution that delivers powerful results across organization
- Full range of algorithms for your business problems
- Big Data enabled (Hadoop, SQL Pushback)

End-to-end solution

- Data preparation through real time interactions
- Use structured, unstructured and semi-structured data
- Integrated portfolio for business analytics
- Scales from a single desktop to an enterprise deployments

Ithaca College in New York uses analytics to make better admissions decisions and increase students' chances for success

32% increase

in applications in first two years after solution implementation

31% growth

in the number of minority students enrolled

3% higher

first-year retention rates as a result of more insightful admissions processes

Solution components

- IBM® SPSS® Modeler V15
- IBM SPSS Statistics
- IBM Training



Business challenge: For private, not-for-profit universities such as Ithaca College (IC), assembling a freshman class that will stay through to graduation makes both good economic and academic sense. However, with the number of potential applicants in its region declining even as internal pressure to increase its minority population grew, IC needed to boost the number of applications it received and find a better way to identify the most promising candidates.

The smarter solution: IC's director of enrollment planning, Yuko Mulugetta, and her team turned to analytics for the answer, digging deeply into data about past and present students as well as metrics from its internal social media platform to find correlations that provide administrators with unprecedented insight into the factors that lead to student success. As a result, the college dramatically changed the way it approaches admissions, including making standardized tests optional.

"We have access today to data that we didn't several years ago, along with great tools that help us make better, more informed decisions."

—Eric Maguire, vice president for enrollment and communication

Michigan State University uses predictive analytics to boost fundraising by better aligning programs with actual donor interests

Identifies donors

with as high as an 85% probability of making a sizable gift

Prioritizes

events and programs based on projected donor interest

Strengthens

relationships with patrons through targeted touches built with personalized profiles

Solution components

- IBM® Cognos® Business Intelligence V10
- IBM SPSS® Modeler
- IBM SPSS Modeler Server
- IBM Business Partner The Ironside Group, Inc.



Business challenge: To offset shrinking government support, Michigan State University needed to strengthen its private fundraising efforts by better identifying those events and programs most likely to solicit gifts.

The smarter solution: The university deployed a predictive analytics engine that mines more than 100 million records—ranging from donation histories to social media content to market reports—and evaluates 177 variables to identify donor patterns and assign affinity scores and donation probabilities to contributors. Now university employees can align fundraising programs with donor interest based on real-world data. For example, the organization noticed historical interest in galas featuring the university's president and has thus increased similar events.

“By unlocking the knowledge that has been hidden in our structured and unstructured data ... we have a much greater chance of matching the right benefactor with the right opportunities.”

—Monique Dozier, AVP, advancement information systems and donor strategy



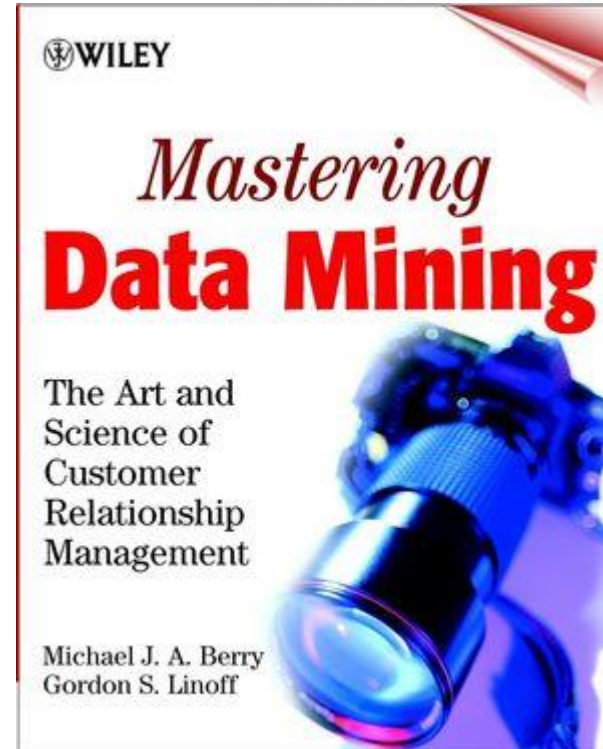
THANK
YOU

The text 'THANK YOU' is rendered in a large, 3D, light blue font. Each letter of the word 'THANK' contains a different portrait of a person. The 'T' shows a man in a white shirt and orange tie. The 'H' shows a woman in a green top. The 'A' shows a man in a green shirt. The 'N' shows a woman in a blue patterned top. The 'K' shows a man with glasses in a blue shirt. The word 'YOU' is positioned below 'THANK'. The 'Y' shows a man in a light blue shirt. The 'O' shows a man in an orange shirt. The 'U' shows a woman in a dark top. The background features a faint, light blue grid pattern in the upper right corner.

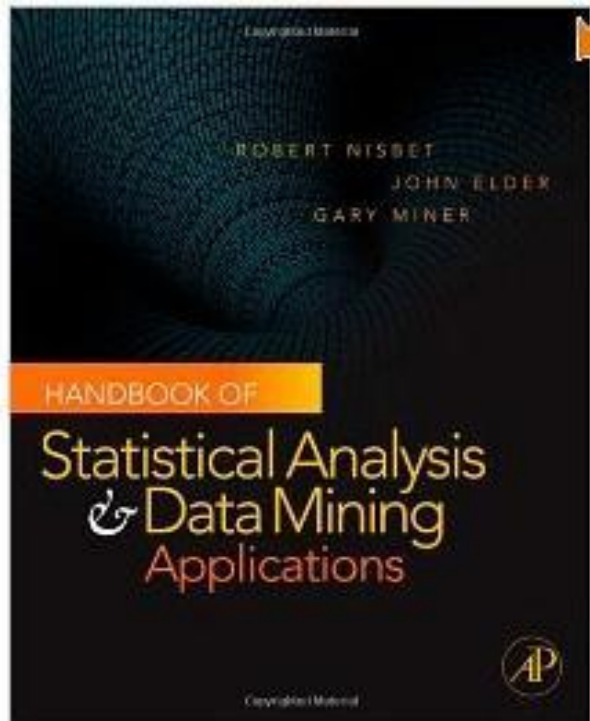
	<h1>Appendix</h1>
--	-------------------

Data Mining Overview

- From Amazon.com
 - Paperback: 512 pages
 - Publisher: Wiley; 1 edition (December 28, 1999)
 - Language: English
 - ISBN-10: 0471331236
 - ISBN-13: 978-0471331230 ;
- Good introductory text on data mining for marketing from two top communicators in the field



Statistical Analysis and Data Mining



- Handbook of Statistical Analysis and Data Mining Applications
- Robert Nisbet, John Elder IV, and Gary Miner
- Academic Press (2009)
- ISBN-10: 0123747651

- An excellent guide to many aspects of data mining including Text mining.

Data Mining Algorithms

- From Amazon.com
 - **Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations**
 - by Eibe Frank, Ian H. Witten
 - Paperback - 416 pages (October 13, 1999)
 - Morgan Kaufmann Publishers;
 - ISBN: 1558605525;
- **Best book I've found in between highly technical and introductory books. Good coverage of topics, especially trees and rules, but no neural networks.**

