Quantitative Models for Predicting Project Success

Philip Paul, Unlimited Innovation, Inc.

Dean Caccavo, Rick Hefner, Kevin Schaaff, Diane Miller
Northrop Grumman Corporation

CMMI Technology Conference & User Group
November 12-15, 2007

Background

- **Predictive modeling is an essential skill at CMMI Levels 4 and 5**
  - Organizational Process Performance requires predictions based on statistical analysis of the organization's standard process
  - Quantitative Project Management requires predictions based on statistical analysis of the project's defined process

- **Predictive modeling relies on historical program performance data (predictive analytics) in conjunction with a forecasting algorithm model to predict future outcomes**
  - Ranges from simple extrapolation techniques to sophisticated Neural Network based models

- **This presentation will discuss the principles of predictive modeling, outline the fundamental methods and tools, and present typical results from applying these techniques to project performance**

Empowering Your Enterprise

Agenda

- ✓ **What is Predictive Analysis?**
- ▪ **Recent Trends**
- ▪ **Application to Program Performance**
- ▪ **Summary**

3

Empowering Your Enterprise

What is Predictive Analysis?

- *Could this network packet be from a virus attack?*
  - Predict likelihood of the network packet pattern
  - → **Anomaly detection (outlier detection)**
  - Similar questions:
    - Are the hospital lab results normal (Adverse drug effect detection)
    - Is this credit transaction fraudulent? (fraud detection)

- *Will this student go to college?*
  - Based on Gender, ParentIncome, ParentEncouragement, IQ, etc.
  - E.g., if ParentEncouragement=Yes and IQ>100, College=Yes
  - → **Classification (prediction)**
  - Similar questions:
    - Is this a spam email? (spam filtering)
    - Recognition of hand-written letters (pen recognition)

- *What is the person's age?*
  - Based on Hobby, MaritalStatus, NumberOfChildren, Income, HouseOwnership, NumberOfCars, …
  - E.g., If MaritalStatus=Yes, Age = 20+4*NumberOfChildren+0.0001*Income+…
  - → **Regression (prediction)**

Agenda

- **What is Predictive Analysis?**
- ✓ **Recent Trends**
- **Application to Program Performance**
- **Summary**

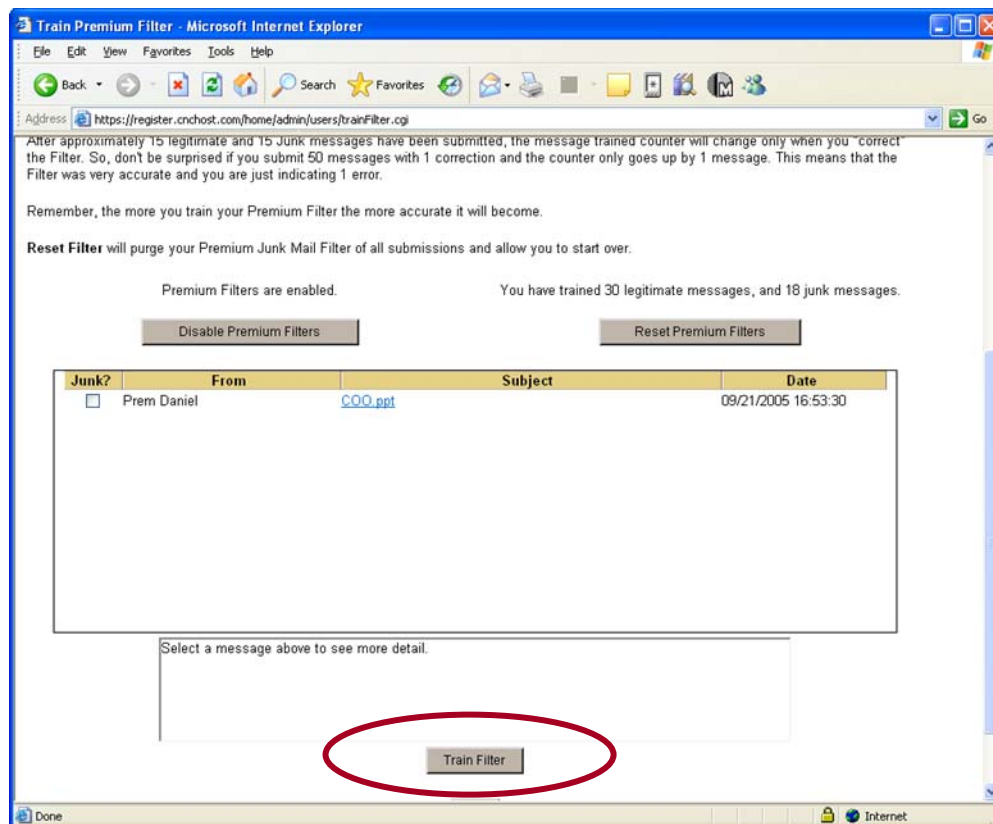**Empowering Your Enterprise**

Predictive Analysis Trends – Adoption is on the rise

- Predictive Analysis is becoming more prevalent and integrated in business applications
  - o *Example: Disease management and evidence based care, based on historical diagnosis and procedure codes of patients*
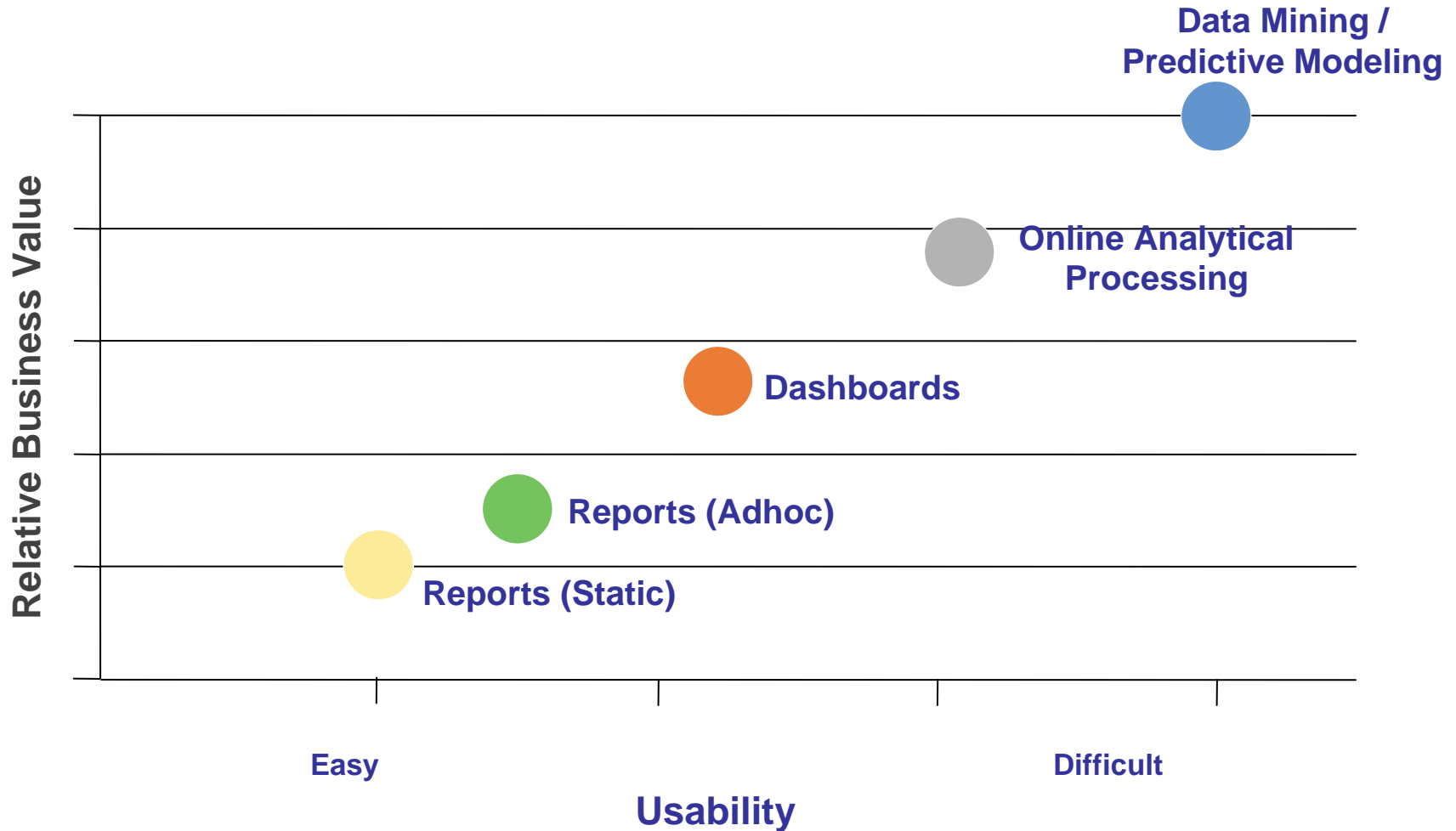 - o *Example: E-Mail filtering using predictive analysis*

- Predictive Analysis algorithms are being integrated into existing databases, data mining tools
  - o *Example: Microsoft SQL Server 2005 has predictive analysis algorithms*
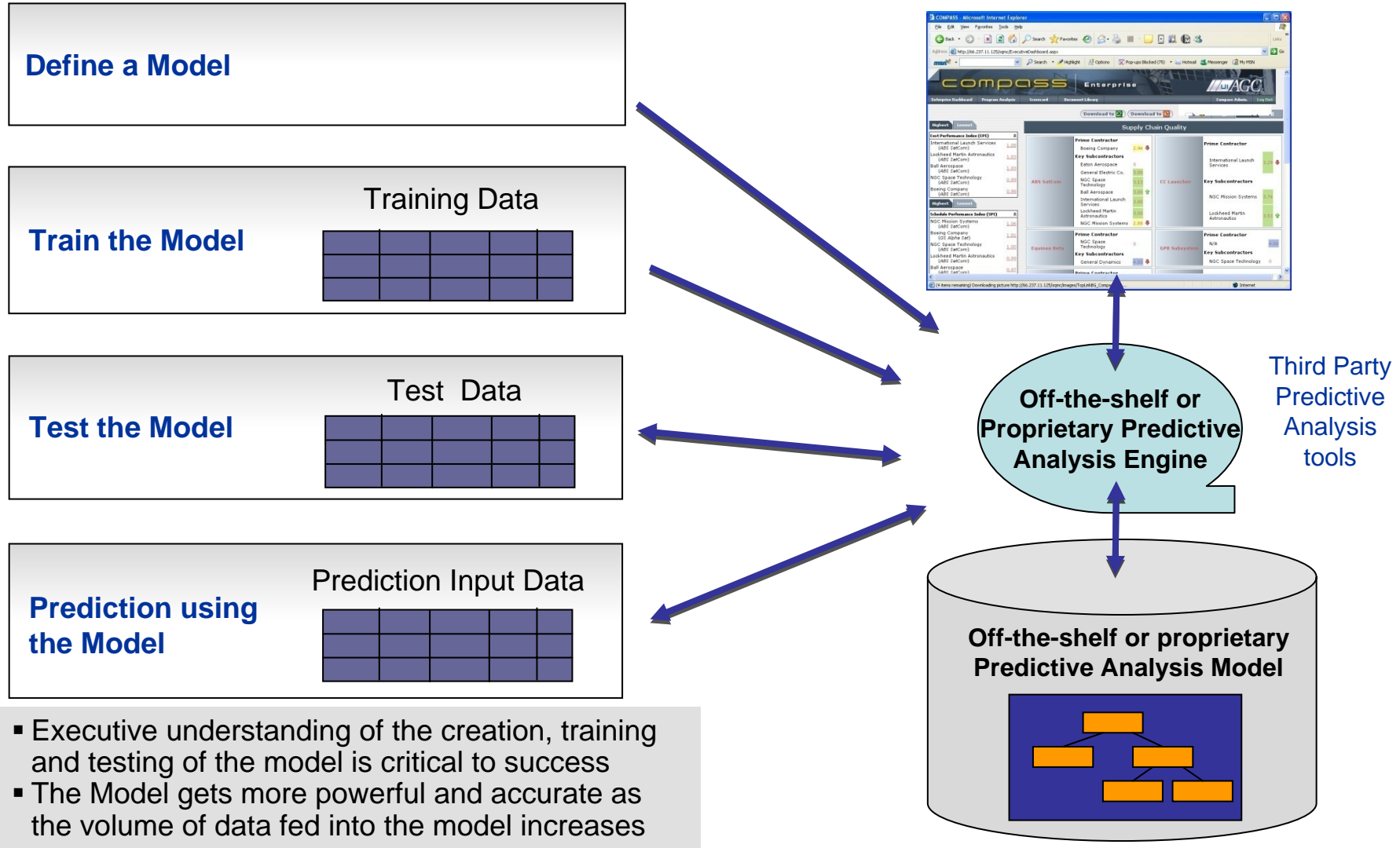
Example:
Premium predictive analysis based filtering on e-mail, available to any e-mail user

Empowering Your Enterprise

Predictive Analysis Trends – Tools are becoming easier to use

**Data Mining /
Predictive Modeling**

Relative Business Value

**Online Analytical
Processing**

Dashboards

Reports (Adhoc)

Reports (Static)

Easy

Difficult

Usability

Empowering Your Enterprise

Predictive Analysis Trends – Model development is more structured

Define a Model

Train the Model

Training Data

Test the Model

Test Data

Prediction using the Model

Prediction Input Data

Off-the-shelf or Proprietary Predictive Analysis Engine

Third Party Predictive Analysis tools

Off-the-shelf or proprietary Predictive Analysis Model

- Executive understanding of the creation, training and testing of the model is critical to success
- The Model gets more powerful and accurate as the volume of data fed into the model increases

Predictive Analysis Trends – Algorithms are available for use

Decision Trees	Naïve Bayesian	Clustering	Sequential Clustering	Time Series	Association rules	Neural Network	
1	2	2	2		2	1	Classification
1	2	2	2			1	Regression
		1	1			2	Segmentation
1	1	2	2		1	1	Association Analysis
		1	1			2	Anomaly Detect.
			1				Sequential Analysis
				1			Time series

1 - First Choice   2 - Second Choice

Empowering Your Enterprise

Data Mining Vendors & Tools

- SAS (Enterprise Miner)
- IBM (DB2 Intelligent Miner)
- Oracle (ODM option to Oracle 10g)
- SPSS (Clementine)
- Insightful (Insightful Miner)
- KXEN (Analytic Framework)
- Prudsys (Discoverer and its family)
- Microsoft (SQL Server 2005)
- Angoss (KnowledgeServer and its family)
- DBMiner (DBMiner)
- Many others

Agenda

- **What is Predictive Analysis?**
- **Recent Trends**
- ✓ **Application to Program Performance**
- **Summary**

Empowering Your Enterprise

Mission Assurance Continuum

Program Performance Oversight	Program Analysis Reporting	Predictive Program Health
Industry Minimum	**Industry Best Practice**	**Industry Innovators**
Proactive Program Management Program Portfolio Management	Reports based on current and passed performance data of portfolio programs, programs, and subcontract reports	Predictive Analysis based on Program Performance Modeling

Approach and Scope

• Self reported Program Portfolio includes critical and high visibility programs • Standard Program Management Metrics collected on a periodic basis	• Self Reported Program metrics collected periodically and at specific program milestones • Reporting analysis performed as needed	• Self reported program metrics, organizational data, personnel data and customer reported metrics collected at regular intervals • Predictive models developed using historical data (leading indicators rationalized) • Models validated against historical data

Infrastructure and Breadth

• Program data maintained by individual programs • Summary information provided to enterprise repository	• Program data collected periodically into an enterprise-wide program management repository • Program, Enterprise and Subcontracts performance reports available	• Holistic enterprise wide approach to program execution • Models continually refined using current program performance data • Sophisticated predictive measures provided to programs and enterprise

Data Requirements

|		
• Very few metrics collected from programs • Key program metrics (cost performance, schedule performance, technical performance, CPI, SPI etc.) • Standardized program taxonomy information like customer, contract type etc.	• 25 – 100 metrics collected from programs • Key program metrics collected at all specified Program Milestones.	• 50 – 75 metrics collected from programs and refined to include only the few relevant metrics • Adaptive approach to qualitative and quantitative performance indicators • Direct and Indirect metrics collected for the programs; qualitative information is mined • Proactive responses based on predictive analysis of ongoing and historical performance

Overarching Objectives for Predictive Modeling

- **Provide program management staff with Predictive Models to "test-their-gut" against enterprise experience data before making strategic program decisions**

- **Develop Predictive Models that provide insight into identifying "headlight metrics" that influence Schedule and Cost realism during program execution**

- **Leverage existing enterprise information to develop Predictive Models for programs**

- **Ensure that models are extensible and automatically calibrated with additional data from the program and enterprise**

Empowering Your Enterprise

Potential Areas for Predictive Analysis

Potential Predictive Analysis Models for Program Management and Subcontractor Management

- Schedule Risk at WBS level based on past performance
- Cost Risk at WBS level based on past performance
- Technical Risk at WBS level based on past performance
- Spending and staffing profile for the program life cycle
- Subcontractor risk profile based on past performance
- Sub-tier quality at subcontract and WBS level
- Defect/Aberrations for the program life cycle
- Mission Assurance models based on program category
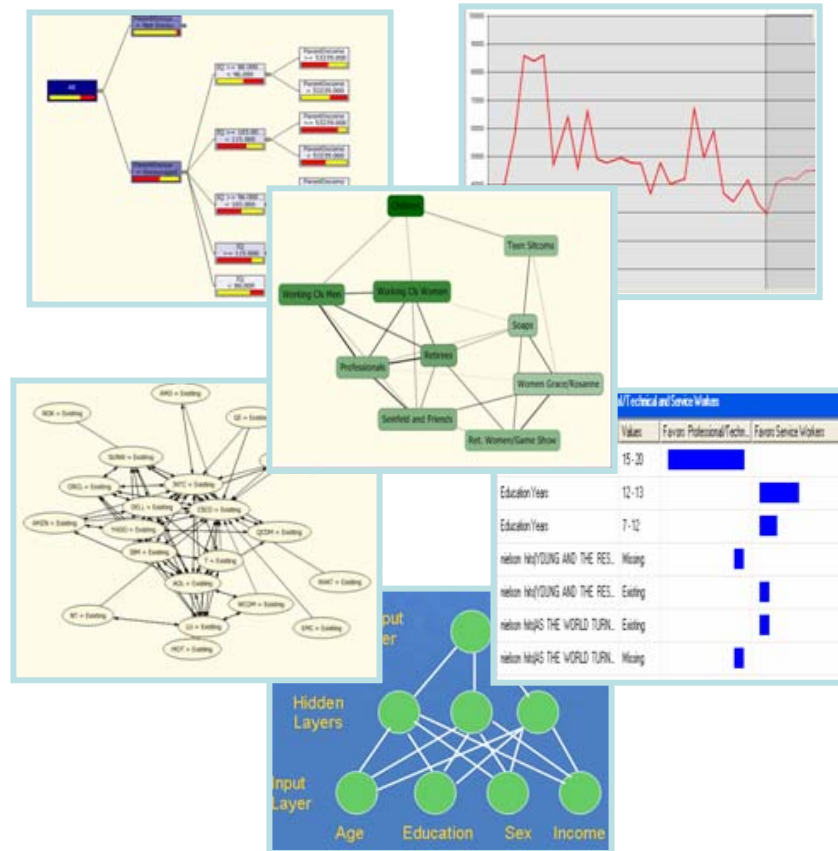
Predictive Analysis Algorithms

- Decision Trees
- Naïve Bayesian
- Clustering
- Sequence Clustering
- Association Rules
- Neural Network
- Time Series
- Custom Model

Empowering Your Enterprise

Predictive Analysis High Level CONOPS

1) Enterprise data is mined and analyzed

2) Enterprise models are defined by Analysts

3) Enterprise model outputs are defined by Analysts and customized by PM staff

4) PM staff use models interactively

Key Benefit:
Leverages enterprise experience data and sophisticated algorithms into predictive models for cost and schedule realism checks during program execution

Program Management Staff Model output review 'What If' Analysis

Predictive Analysis Output Pivot Chart

Predictive Analysis Output Lift Graph

Predictive Analysis Output Cluster Diagram

Predictive Model Engine

Enterprise Models (Customer)

Enterprise Models (Schedule)

Enterprise Models (Technical)

Text with Data Repository

Reporting Feeds | Historical Feeds | Project Management Feeds | EVMS Feeds

Manual Input | Historical | PM Systems | EVM Systems

Empowering Your Enterprise

The Predictive Modeling Process

- **Explore the Data**
- **Understand Data Relationships**
- **Derive/Enhance the Data**
- **Use the Data to Predict**
- **Train the Model**

Empowering Your Enterprise
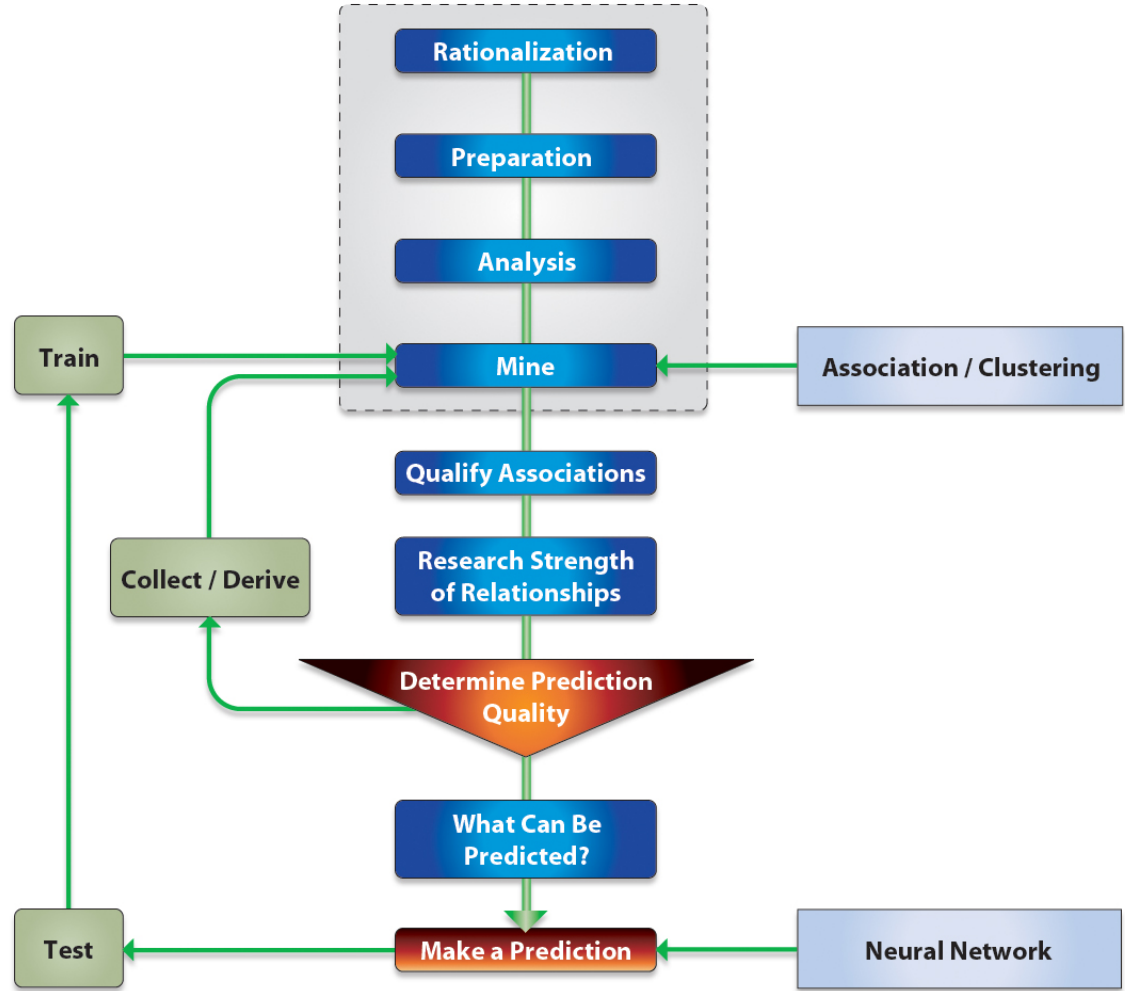
What can be Predicted with Reasonable Accuracy?

Limited Number of Programs

Enterprise Experience

Program Lifecycle Stage

Large volume of historical data

- Likelihood or return to acceptable performance
- Predictive Program Performance

1

- Quadrant 2 predictions
- Quadrant 3 predictions
- Early warning "headlight indicators"
- Higher accuracy based on enterprise experience

3

Limited Historical data

- Cost, schedule realism
- Phase realism
- WBS Accuracy

2

Low

High

Volume of "Like" Programs

Derivation of Data & Data Relationships

- **Examples of Derived Data**
  - Number of Outstanding Program Issues (with and without recovery dates)
  - Variance in program Cost/Schedule/Technical health from month-to-month
  - Program Cost/Schedule/Technical health trend from month-to-month
  - Variance in VAC from month-to-month taken as a percentage of the current EAC

- **Examples of Discovered Relationships**
  - Schedule Health is a good indicator of program Overall Health recovery
  - Cost and Technical Health are good indicators of program Overall Health decline

Better understanding of the data allows for organization and enhancement of the dataset

Empowering Your Enterprise

Model Development & Calibration

Model	**Calibrated Model**
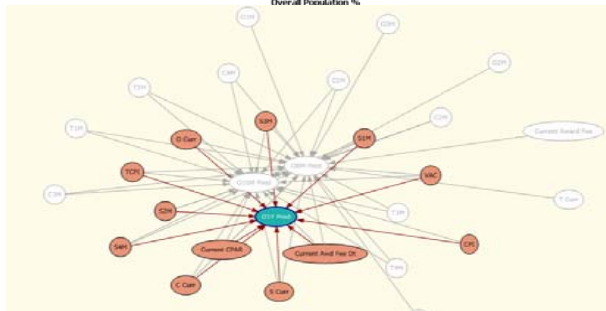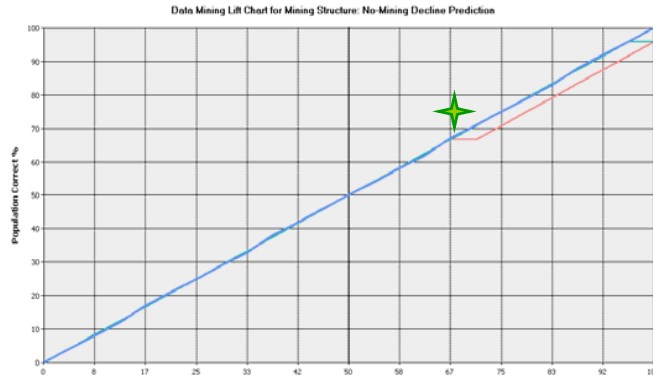
- Modeling without applied domain knowledge or calibration resulted in lower accuracy
- Association models able to determine relevant data attributes
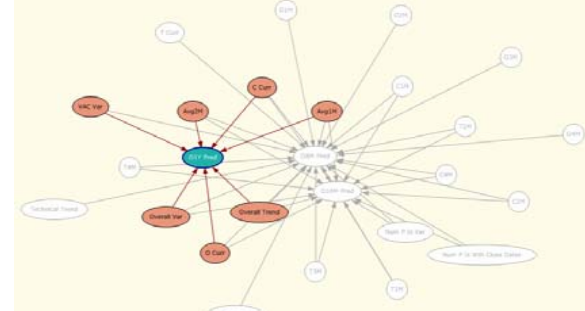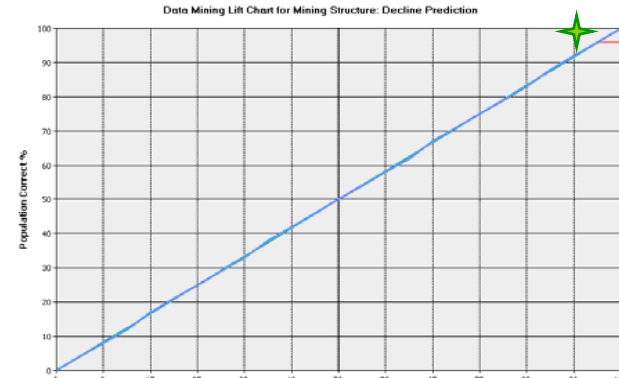
- Incorporating domain knowledge and calibration into data mining resulted in higher accuracy
- Data relationships are more clearly defined

Domain knowledge & calibration applied to data mining can enhance the predictive model

Empowering Your Enterprise

Presentation of the Results from the Models

Ability for Programs to review the predictive output from multiple models to "test-the-gut" before making strategic program decisions

FICTIONAL DATA

Empowering Your Enterprise
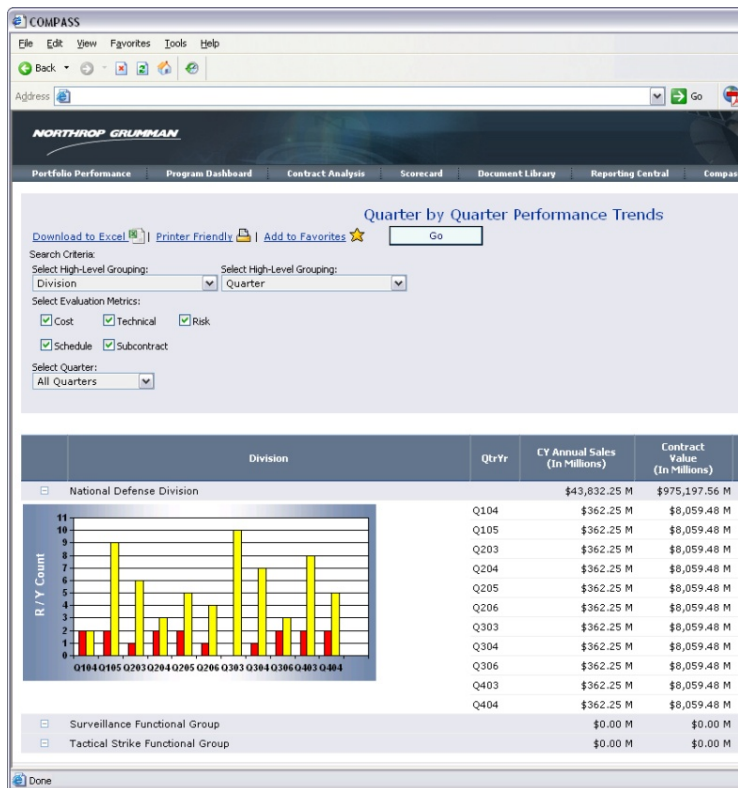
Presentation of the Results from the Models

Ability for staff to review status and trends across the portfolio of programs, across a variety of categories

FICTIONAL DATA

Algorithm References

- **Decision trees (classification/regression):**
    - ftp://ftp.research.microsoft.com/users/surajitc/icde99.pdf
    - http://www.research.microsoft.com/research/pubs/view.aspx?tr_id=81
    - http://research.microsoft.com/~dmax/publications/dmart-final.pdf

- **Association rules:**
    - Apriori algorithm (see Data Mining concepts and techniques)
- **Clustering**
    - EM:http://www.research.microsoft.com/scripts/pubs/view.asp?TR_ID=MSR-TR-98-35
    - K-means (see Data Mining concepts and techniques)

- **Sequence clustering**
    - ftp://ftp.research.microsoft.com/pub/tr/tr-2000-18.pdf

- **Time series:**
    - http://research.microsoft.com/~dmax/publications/dmart-final.pdf

- **Neural network**
    - Conjugate gradient method (see Data Mining concepts and techniques)
- **Naïve Bayesian**
    - See Data Mining concepts and techniques

More Information

- OLE DB for DM specification
  - http://www.microsoft.com/downloads/details.aspx?FamilyID=01005f92-dba1-4fa4-8ba0-af6a19d30217&DisplayLang=en
- Plug-in
  - http://www.msnusers.com/AnalysisServicesDataMining/Documents/Files%2FSQL%20Server%20Data%20Mining%20Plug%2DIn%20Algorithms%20%28Beta%202%20%2B%2B%29.zip
  - A white paper, tutorial, and complete sample code for Pair-wise Linear Regression
- SQL Server 2005:
  - www.microsoft.com/sql/2005
- Community:
  - Microsoft.public.sqlserver.datamining
  - Microsoft.private.sqlserver2005.analysisservices.datamining
  - Groups.msn.com/AnalysisServicesDataMining
- msdn.microsoft.com (search "data mining")

Agenda

- **What is Predictive Analysis?**
- **Recent Trends**
- **Application to Program Performance**
- ✓ **Summary**

Empowering Your Enterprise

Summary – Critical success factors

- Executive and Enterprise support and understanding of long-term strategic benefits
- Understanding of the types of data and the correlation between the data
- Understanding of the various constituents in the value chain and the tools/processes for each constituent
- Prototypes or mockups that depict the results of the model
- Sound and robust technical architecture
- Delivery mechanism that shields the complexity of the model from the end users

25

Empowering Your Enterprise

Contact Information

Philip Paul
Unlimited Innovations, Inc.
(714) 342-8952
ppaul@uius.com

Rick Hefner, Ph.D.
Northrop Grumman Corporation
(310) 812-7290
rick.hefner@ngc.com

Empowering Your Enterprise