

Predictive Analytics

University of Colorado Boulder - November 12, 2013



Michael Walker

Managing Partner

720.373.2200

m@rosebt.com

<http://www.rosebt.com>



Fortune Teller



Aristotle - Natural Philosophy



Galileo Tower of Pisa Experiment



Francis Bacon - Scientific Method



Scientific Method

Francis Bacon - Novum Organum, 1620

Narrow scope of natural philosophy practiced

Inductive Logic

Goal is improved engineering - not philosophical truth

Basic and applied research

Scientific Method

Careful experimentation + theorizing

Science allows us to make predictions

Test theories with well designed experiments

Karl Popper



Scientific Method

1. Falsification - design test to prove theory wrong

Scientific statement must be nonobvious, falsifiable
predictive rule

2. Science never provides the truth - may fail future
test

3. Theory precedes experiment

4. Replication

Scientific Method

Science does not tell us if theories are true

Only if they are true in the sense of allowing us to make reliable, nonobvious predictions

Thus, science is about predictions

Garden of Eden vs. Black Box

Garden of Eden - cause & effect or deterministic

If only we had the vast knowledge of God, everything could be understood and predicted

If only you have enough good data and compute power, you can make accurate predictions

Garden of Eden vs. Black Box

Black-box - probability theory - stochastic view

We can see what goes into the box and what comes out, but not what happens inside

We can only draw inferences about the odds of input A producing output B

We cannot follow the path of every molecule in gas, but we can work out its average energy and probable behavior and design a useful pipeline to transport gas

Causality

Implicit assumption that if you know the cause, you can forecast events and manage risk

Real world is not that simple - causes are usually obscure

Critical information often unknown or unknowable

Causes can be concealed or misrepresented

High causal density environments

Probability Theory

Analysis of random phenomena - random variables, stochastic processes, and events: mathematical abstractions of non-deterministic events or measured quantities that may either be single occurrences or evolve over time in an apparently random fashion

Need to learn to think probabilistically

Events are not perfectly predictable but can be described by mathematical laws of chance

Risk is measurable and manageable - up to a point

Probability Theory

Bayesian probability

Frequentist probability

Algorithmic probability

Pignistic probability

Inverse probability

Knightian uncertainty

Fuzzy logic

Bernoulli stochastics - ignorance space

Propensity probability

Calculus of predispositions

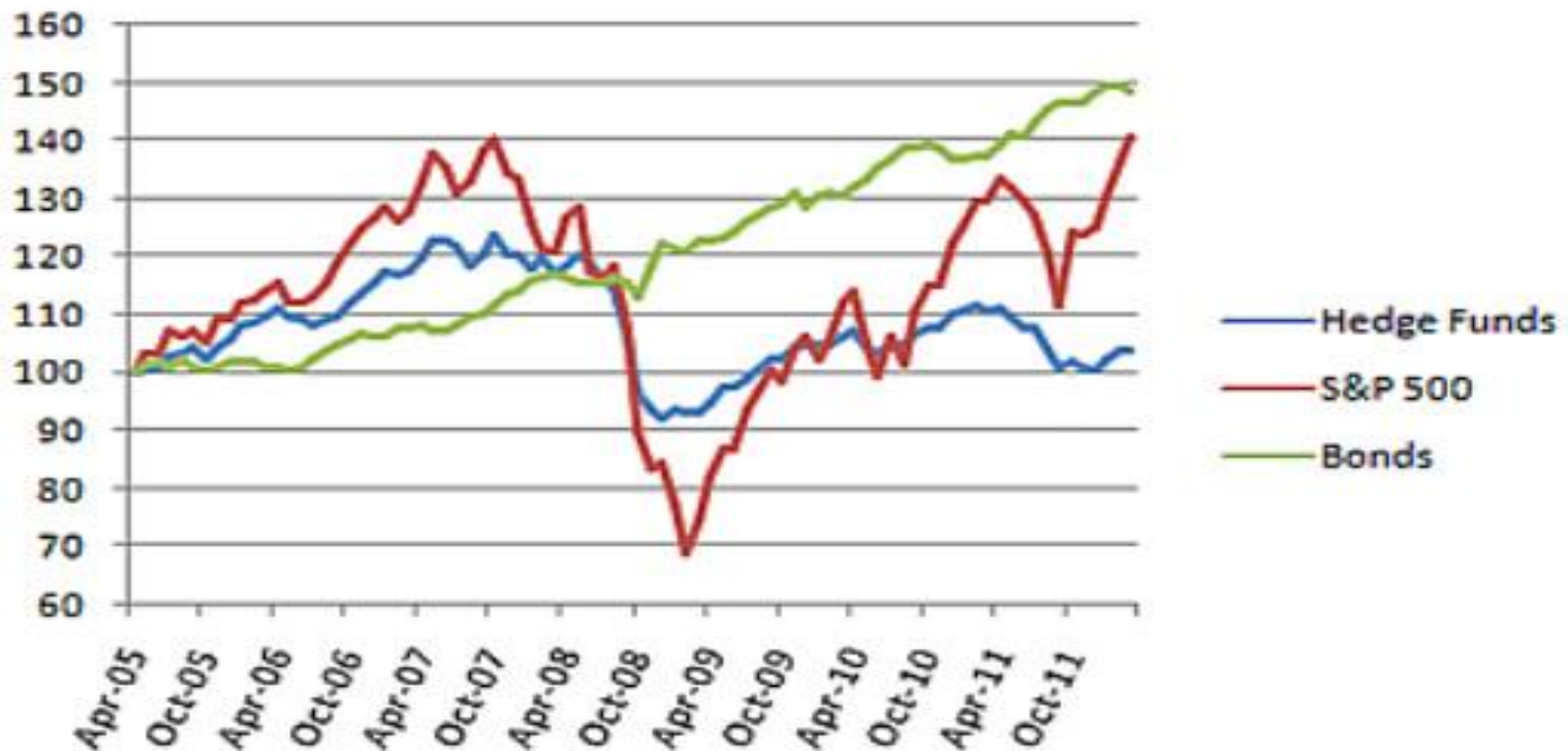
Finance

Can silicon intelligence find profitable patterns that humans cannot?

Computational techniques: algorithms, neural networks, etc.

To date, this has failed.

High causal density environment.



If you had put your money into hedge funds in 2005 and distributed it across the whole universe of hedge funds, you would have roughly the same amount of money today.

By contrast, if you had simply bought the existing universe of publicly traded stocks and bonds and reinvested interest and dividends, your portfolio would have grown by half.

Finance

Few authentic geniuses in the hedge fund business who throw off spectacular returns regularly.

Larger number of clever dudes who have a great idea one year (like John Paulson, who foresaw the housing collapse) and then lose just as spectacularly the next year.

And there are a very large number of Ivy-educated herd-followers in pink shirts and suspenders with no particularly notion of what they should do.

Finance

All use big data + compute power + quantitative analysis.

Army of finance data scientists (quants).

Rely mostly on post-modern finance theory, models and algorithmic trading.

Nobody has hit jackpot.

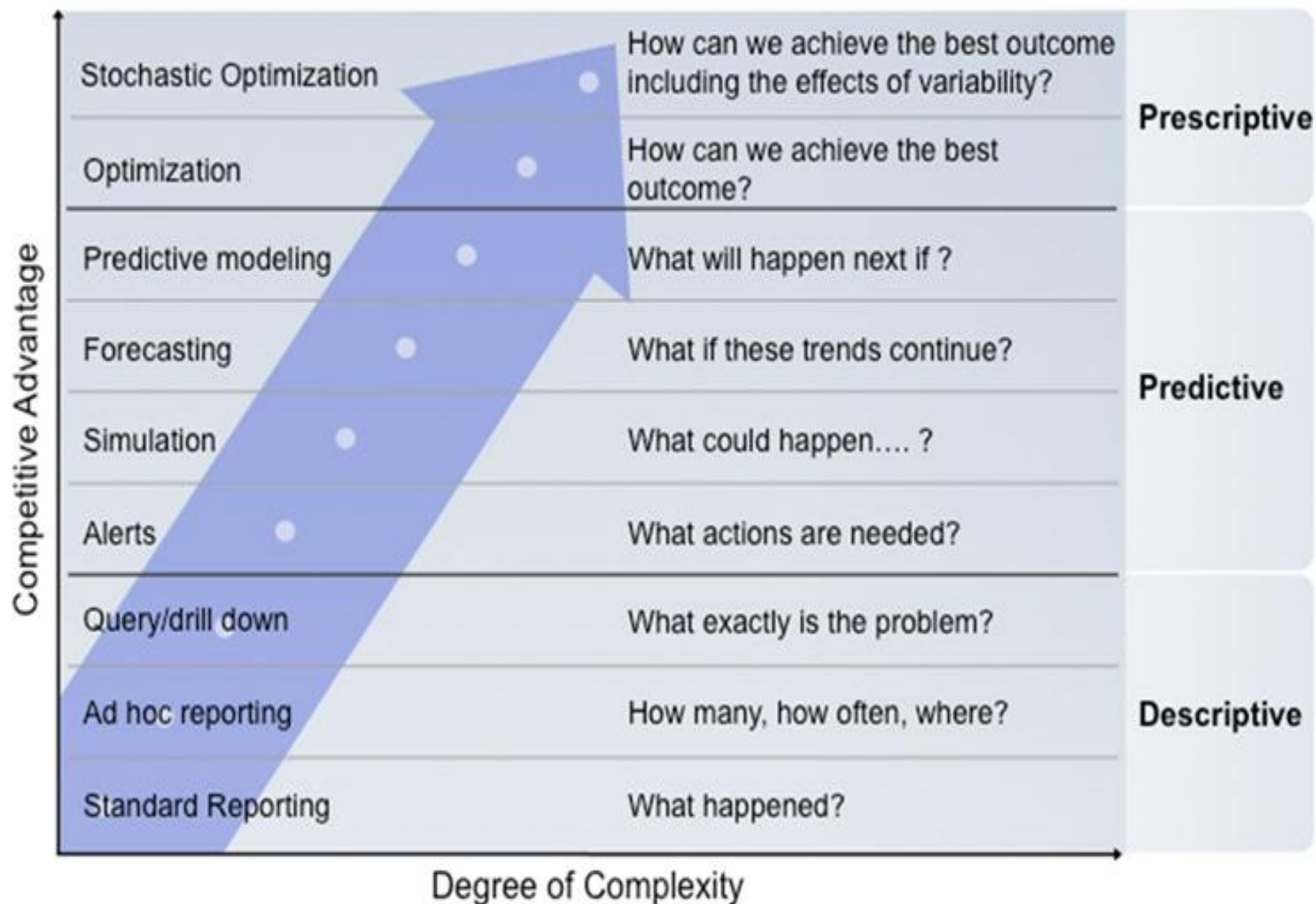
Predictive Success Formula

Humans + computers outperform humans alone or computers alone.

Experiment - freestyle chess competitions.

Successful traders use data science, HPC and brainpower.

They see what others do not see and experiment. Use models but do not rely on them for making decisions.



Based on: Competing on Analytics, Davenport and Harris, 2007

Is your organization trying to find new ways to generate revenue?

Most organizations apply predictive analytics to core functions that produce revenue.



52%

use predictive analytics to increase profitability

55%

use predictive analytics to create new revenue opportunity

Are you looking for ways to increase customer satisfaction?

45%

of organizations currently use predictive analytics for customer services



What data in your organization can predictive analytics tap to help you discover new trends and opportunities?

The **top 5** sources of data tapped for predictive analytics:



40%

of companies are evaluating or plan to use social media data

Are you using the right product offers and targeting the best customers for your up-sell and cross-sell?

43%

of organizations currently use the results of predictive analytics for product recommendations and offers



What impact can predictive analytics have on your organization?

86% assert that predictive analytics will have a **major positive impact** on their organization.

with nearly **one-third** indicating it could be transformative in enabling them to do things they couldn't do before



How can your organization gain a competitive edge and respond in real time?

68% of organizations who use predictive analytics have **realized a competitive advantage**.

With **real-time predictive analytics**, you can make sure your company doesn't miss its window of opportunity.



What are the 5 things predictive analytics can do for you?

- 1 Instantly predict market trends and customer needs
- 2 Create customized offers for each segment and channel
- 3 Predict how market-price volatility will impact your production plans
- 4 Forsee changes in demand and supply across your entire supply chain
- 5 Proactively manage your workforce by attracting and retaining talent

Where do you want your company analytics to be?

Predictive analysis enables you to extend your analytics capabilities: Moving from the rearview mirror to a **forward-looking view**.

What happened?

Standard reports
Ad Hoc reports
OLAP analysis

Why did it happen?

Drill down analysis
Data discovery

What will happen?

Predictive modeling

What is the best that could happen?

Real-time predictive analysis

Sense and Respond

Predict and Act



Predictive Analytics

Three basic cornerstones:

Predictive Modeling

Decision Analysis and Optimization

Transaction Profiling

Predictive Analytics Tools

Scientific methods

Experiments

Analytical techniques

Machine learning techniques

Algorithm design and execution

Data visualization and story-telling

Statistics

Math

Computer engineering

Data mining

Data modeling

Predictive Analytics Techniques

Regression techniques

Linear regression models

Discrete choice models

Logistic regressions

Multinomial logistic regressions

Probit regressions

Time series models

Survival or duration analysis

Classification and regression trees

Multivariate adaptive regression splines

Singular value decomposition

Machine Learning Techniques

Neural networks

Radial basis functions

Support vector machines

Naïve bayes models

K-nearest neighbour algorithms

Geospatial predictive modeling

Practitioner Tools

Bayesian Modeling

Monte-Carlo Simulations

Regression Analysis

Random Forests Algorithm

Good Quality Data

Forecasting Principles

If policy A is adopted then X will occur.

Often forecasts are made for future values of a time-series; for example, the number of babies that will be born in a year, or the likely demand for compact cars.

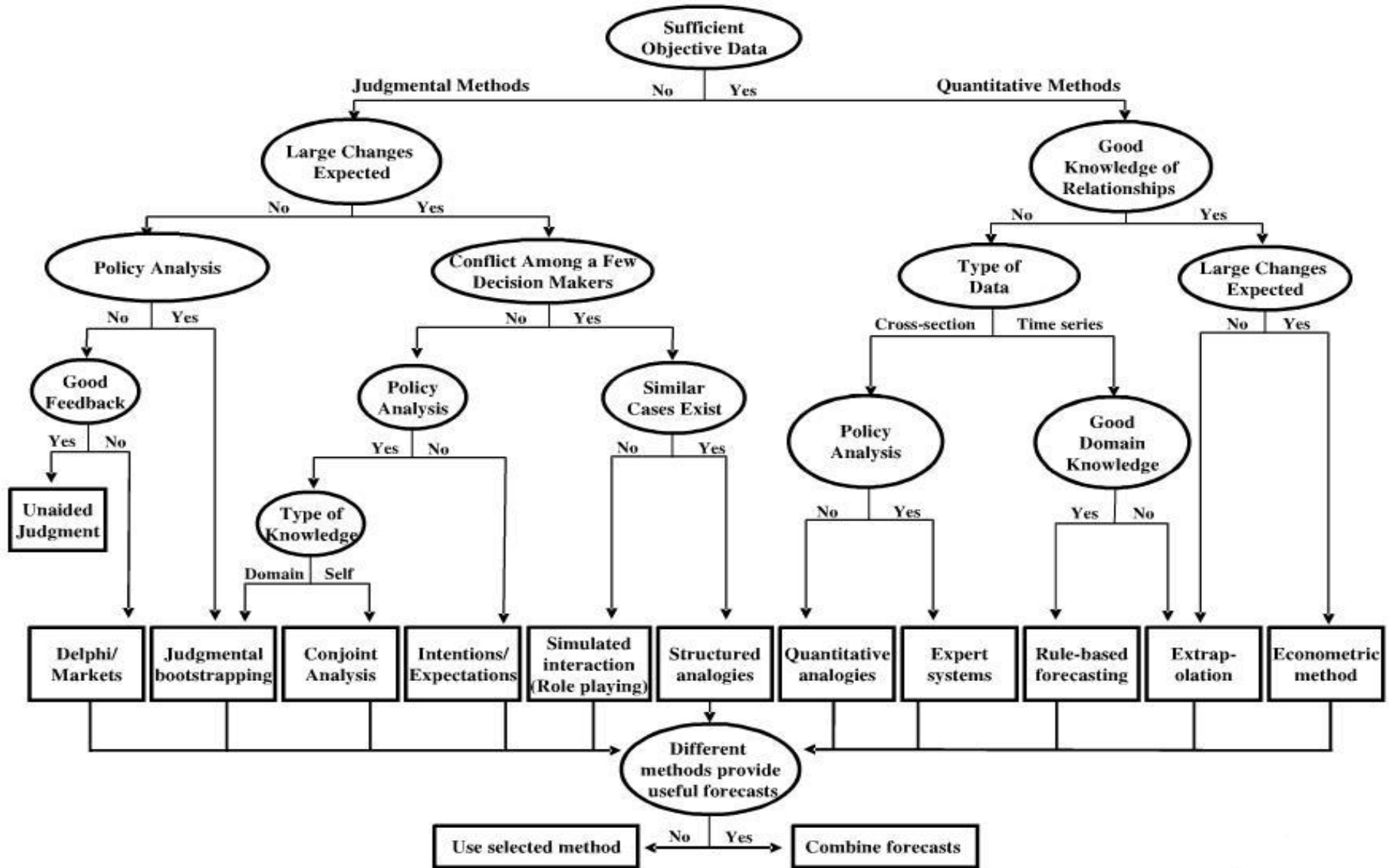
Forecasting Principles

Forecasts can be of one-off events such as the outcome of a union-management dispute or the performance of a new recruit.

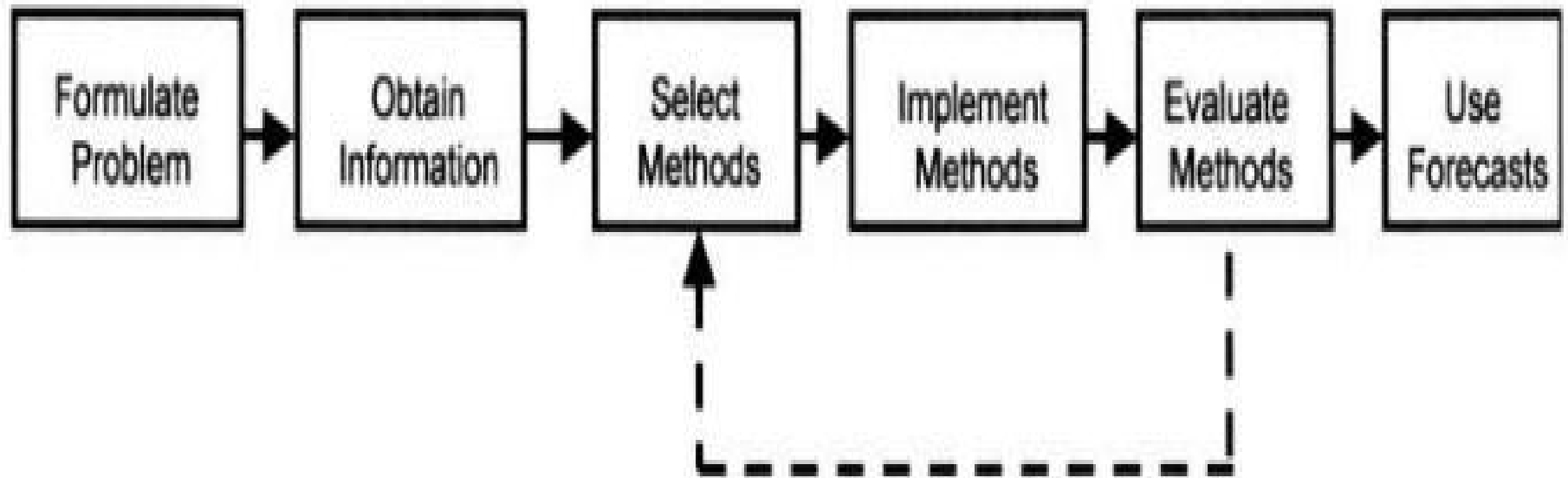
Forecasts can also be of distributions such as the locations of terrorist attacks or the occurrence of heart attacks among different age cohorts.

Forecasting includes the study and application of human judgment as well as of quantitative or statistical methods.

Forecasting Methods Selection Chart



Stages of Forecasting



Forecasting 10 Steps

1. Problem Formation
2. Data Selection
3. Data Preparation
4. Data Exploration
5. Model Building
6. Model Validation
7. Model Deployment
8. Model Evaluation
9. Experimentation
10. Model Iteration

Use Cases - Domains

Retail sales and merchandising analytics [markdown and assortment planning]

Financial services [risk and loan credit scoring]

Pharmaceutical analytics [drug development and clinical trials]

Marketing analytics [CRM, segmentation, and churn analysis]

Text analytics [sentiment analysis]

Financial control analytics [customer payment collections]

Fraud analytics [insurance and medical claims]

Pricing analytics [price sensitivity analysis]

Use Cases - Domains

Telecommunications analytics [customer behavior]

Supply chain and transportation analytics [route optimization]

Manufacturing analytics [warranty claims]

Hospital analytics [patient scheduling]

Human resources analytics [workforce planning]

Banking analytics [anti-money laundering]

Police analytics [crime pattern analytics]

Horizontal & Vertical Applications

- Logistics optimization in the transportation industry
- Price optimization in the retail industry
- Intellectual property management in the media and entertainment industry
- Natural resource exploration in the oil and gas industry
- Warranty management in the manufacturing industry
- Crime prevention and investigation in local law enforcement

Horizontal & Vertical Applications

- Predictive damage assessments in the insurance industry
- Fraud detection in the banking industry
- Patient treatment and fraud detection in the healthcare industry
- Sports strategy
- Human resource management

Use Cases

Predict market trends

Predict customer needs

Create customized offers for each segment and channel

Predict changes in demand and supply across the entire supply chain

Hire the right people

Manage the workforce

Predict who is likely to quit their job

Predict how market-price volatility will impact production plans

Manage risk

Optimizing CRM Systems

Analyze all customer data therefore exposing patterns that predict customer behavior.

With multiple products, predictive analytics can help analyze customers' spending, usage and other behavior, leading to efficient cross sales, or selling additional products to current customers.

Goal: higher profitability per customer and stronger customer relationships.

Customer Profitability

Which customer will generate the most profit from the least effort?

What is the optimal price point for a product or service?

Which demographics are most likely to respond to different types of marketing strategies?

Retail

Which product in a retail store chain can generate the most profit without carrying excess inventory but also not having periods of stock outs?

Fast analysis of buyer patterns, purchase data and call-center communications to understand trends to improve marketing competitiveness, decision-making and profitability.

Detecting and acting on consumer trends and competitors marketing and pricing immediately is critical in retail, especially the online retail space.

Customer Engagement

Understand customers to engage with them in a relevant, timely, personalized manner.

Provide customers with the products and services they want, when and where they want them.

Gain insights into the patterns of customer behavior that relate both positively and negatively to your KPIs and business objectives.

Human Resources

Attracting and Selecting Talent

Hiring the best people for a specific position.

Employee Retention

Which of our employees will be the next most likely to resign and take a job with another company?

Supply Chains

Optimize supply chain performance - forecasting likely product demand.

Data from sensors on trucks or pallets to identify the most optimal delivery route (also taking into account traffic predictions and weather conditions).

Political Campaigns

Nate Silver's model predicted that the probabilities favored Obama.

The University of Colorado Political Science Department - which had accurately predicted each presidential election since 1988 - predicted a Romney win.

Political Campaigns

Model = Aggregate of All Polls

Use more polls, the probable error of the aggregated result decreases.

Factored in more data points than even the most knowledgeable political pro could remember.

Incorporated the size, quality, and time of all polls, and weights them based on the polling firm's past predictive success.

Political Campaigns

Run thousands of daily simulations of possible outcomes using Monte Carlo method. Probability of victory was the percentage of simulations in which a candidate emerged as the winner.

Eliminated the cherry-picking of polls. Set parameters at start, deciding how much weight and accuracy you're going to give to each poll based on historical accuracy. Feed in whatever other conditions you think will matter to the result. Then, sit back and let the algorithm do the work.

Political Campaigns

Team Obama Innovations

Used Data Science & Predictive Analytics

Individual/Group Profiling

Outreach/Message Experiments

Target Unlikely Base Voters to Vote

Marketing

Rank Customers

Predictor = single value measured for each customer.

Recency, based on # of weeks since last purchase -
Higher values for more recent.

Reliable campaign response predictor: more responses from those more highly ranked. Contact customers in order of recency -- first, call the most-recent customer; next, call the next-most-recent customer; and so on -- you will improve your response rate.

Marketing

For each prediction goal, there are an abundance of predictors that will help rank your customer database.

Consider online behavior: Customers who spend less time logged on may be less likely to renew their annual subscription.

Here, retention campaigns can be cost-effectively targeted to customers with a low monthly usage predictor value.

Marketing

Combined Predictors = Better Predictions

Combine two predictors with a formula - simply adding them together. If both recency and personal income influence the chance that a customer will respond to a mailing, a good predictor may be: $\text{recency} + \text{personal income}$.

If recency is twice as important, give it twice the weight: $2 \times \text{recency} + \text{personal income}$.

Marketing

Right combination of predictors will perform better considering multiple aspects of the customer and behavior. To match the complexity of customer decisions, a predictive model must be richer and more complex, combining dozens of predictors.

Trick is to find the best predictive model - there are many kinds of models, such as linear formulas and business rules. Weights or rules determine how predictors combined. So many choices impossible to try them all and find the best one.

Marketing

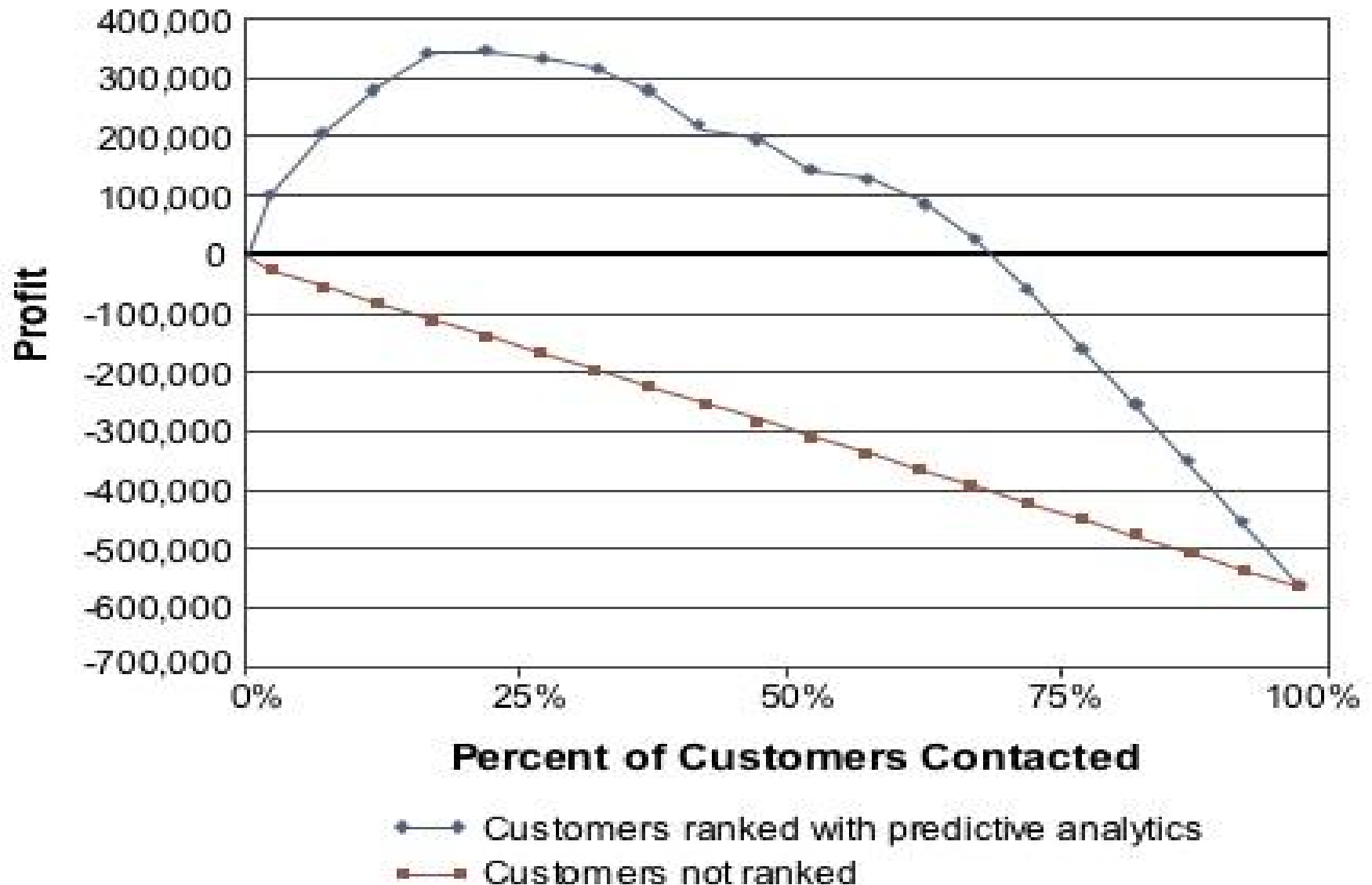
Data mining tech uses customer data to automatically build predictive model - leveraging existing logs of customer purchases, behavior and demographics.

Mixture of number crunching, trial, and error.

Generate simple clear business rules vs. complex formula.

Simpler, intuitive model may not perform prediction as well.

Test Model, Experiment and Continuous Iteration



Prediction Failures

Economic forecast in 2007

NCAR 2006 sunspot forecast

Oregon health insurance experiment

Vitamin E / salt / coffee / cholesterol / fats / carbs

Grades / SAT scores

Class size

Criminal legal theory / 3 strikes law

New coke

Economic stimulus of 2009

Soviet economic strength

Iraq WMD

Prediction Failures

Lesson:

Models flawed yet useful - illusion of reality

Many low-risk experiments vs. few high-risk trials

Beware high causal density environments

Conclusions

Forecasting will improve over time with more and better data and improved machine learning and algorithms.

Integrate nonexperimental (models) and experimental methods.

Be skeptical about assumptions embedded in models.

Many low-risk experiments best strategy.

Conclusions

Business and social science can improve prediction methods by conducting more experiments.

a) use experts

b) experiment design and collection of data vital

Conclusions

Recognition of prediction uncertainty calls for heavy reliance on trial-and-error progress.

Risk management: beware of fat tail events - happen more frequently than statistics predicts.

Conclusions

The limits to the use of trial and error are established predominantly by the need for strategy and long-term vision.

a) no magic prediction methods

b) use probabilistic thinking

c) leadership vital - in short supply

Conclusions

When using a probabilistic forecast, always think:

THE MODEL IS NEVER THE SYSTEM!

Hierarchy of Reliability of Methodologies for Predictive Business Rules

Straight pattern finding for some outcome is least reliable

Pooled regression and other analogous pattern-finding analysis on changes in some outcome is next worse

Quasi-experiments are next worse

True randomized experiments are most reliable

Hierarchy of Reliability of Methodologies for Predictive Business Rules

Competent experts are better than straight pattern-finding models - but not as good as randomized experiments

Randomized experiment is scientific gold standard of certainty of predictive accuracy in business

If program is practically testable and experiment is cost-justified (expected value of info worth cost of test), experimentation dominates all other methods of evaluation and prediction

Better Decisions

Decision makers need a better understanding of the consequences of their decisions - including decisions that may have a profound impact on the business or community.

It is the job of data scientists to give decision makers a range of scientific results of the consequences of different decisions - using probability theory to calculate the odds.

Data Science Classes

The [Data Science Association](http://www.datascienceassn.org) will offer on-line data science classes in January 2014.

Certificate of Data Science Competency

Check site at: <http://www.datascienceassn.org>

Professionalization of Data Science

The [Data Science Association](http://www.datascienceassn.org) - with a membership of over 700 data scientists - has created a [Data Science Code of Professional Conduct](http://www.datascienceassn.org/code-of-conduct.html) and is currently developing data science competency standards.

See: <http://www.datascienceassn.org>

See Code: <http://www.datascienceassn.org/code-of-conduct.html>

Free One Year Membership in the Data Science Association.

Go Join Today: <http://www.datascienceassn.org>

Membership Benefits

Joining an elite professional organization

Valuable credential to signal to clients and employers

Networking access to high-value professionals

Discounts on conferences, books and classes

Voting rights

Thank You

Presentation by:

Michael Walker

Rose Business Technologies

720.373.2200

m@rosebt.com

<http://www.rosebt.com>

