



Advanced Analytics

The Next Wave for Business Intelligence

Kingshuk Banerjee, D.Sc.
Associate Partner, Business Analytics and Optimization
IBM Global Business Services

Setting the Context

Persistence, Business Intelligence and Analytics

3-Layer Cake View



1. Foundational: Data Persistence, data sourcing
Storing or sourcing the data; Enterprise Information Repository, Data-warehouse, Data feeds

Setting the Context

Persistence, Business Intelligence and Analytics

3-Layer Cake View



2. Intermediate: Business Intelligence

Informational; making meaning out of data; on-demand slicing and dicing; roll-up, drill-down real-time

1. Foundational: Data Persistence, data sourcing

Storing or sourcing the data; Enterprise Information Repository, Data-warehouse, Data feeds

Setting the Context

Persistence, Business Intelligence and Analytics

3-Layer Cake View



3. Advanced: Predictive and Prescriptive Analytics

Discovering the Unknown; drawing insights; data mining, text mining, optimization, simulation

2. Intermediate: Business Intelligence

Informational; making meaning out of data; on-demand slicing and dicing; roll-up, drill-down real-time

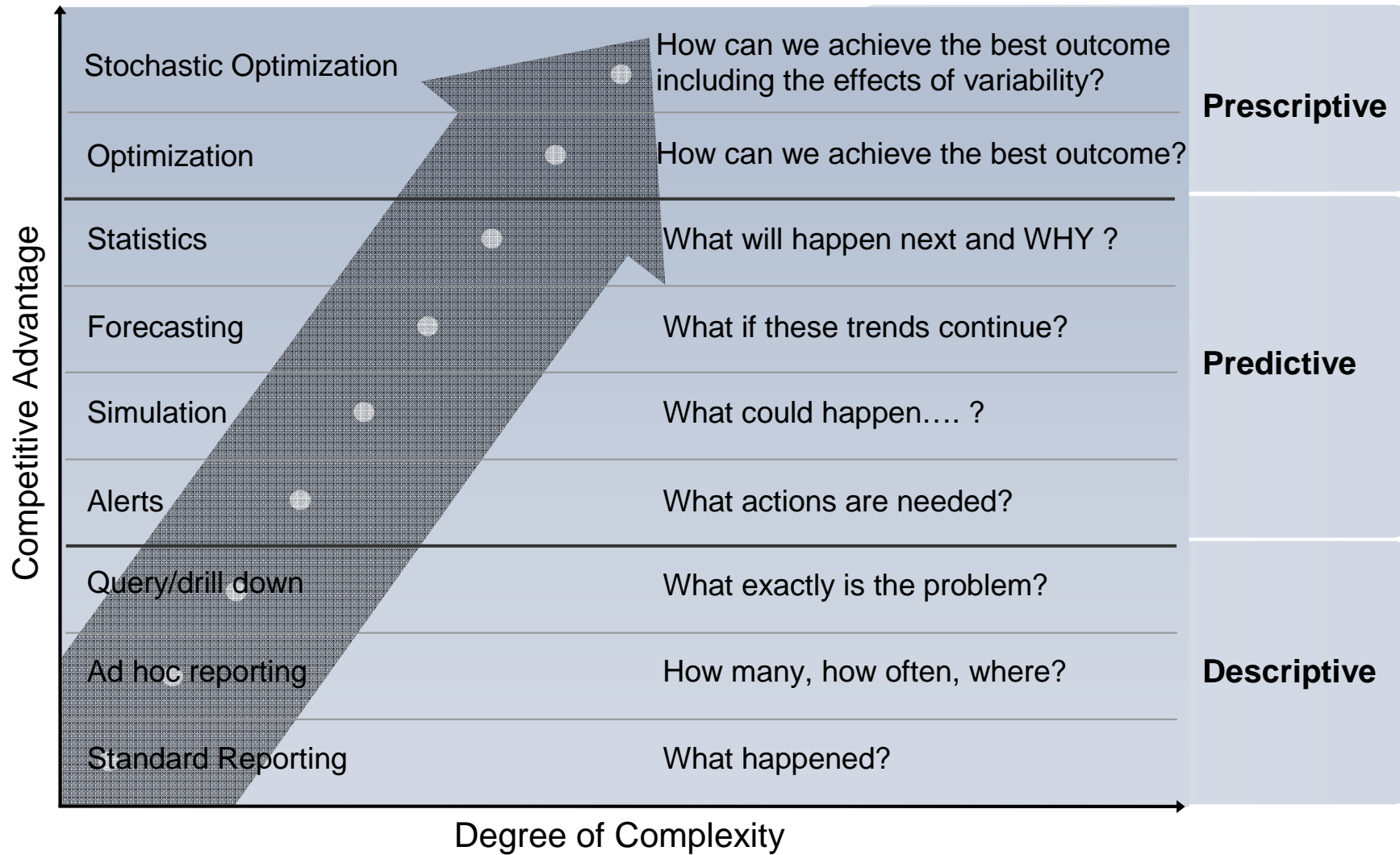
1. Foundational: Data Persistence, data sourcing

Storing or sourcing the data; Enterprise Information Repository, Data-warehouse, Data feeds

Let us Storyboard

- You are the CEO of a Pharmaceutical company
- You want to look into the plummeting sales numbers for your eastern region operation
- You use Business Intelligence tools and techniques to slice and dice the underlying sales and operational data for eastern region to pinpoint the specific drug stores and types of medicine that are selling slower
- You have identified the potential trouble spots, but now you would like to investigate further
 - what is the root cause for this troubling sales trend
 - whether this trend is long-term
 - how to optimize your inventory and supply-chain operations in view of the emerging situation
- Advanced Analytics may be able to help you in above situation
 - Let us briefly discuss how it is different from conventional techniques say Structured Query language (SQL); in SQL you form the query and query-constraints based on what you know; in Advanced Analytics you have no idea of what you are looking for; you expect the smart system tell you what you should know

Analytics – Maturity Levels



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

Advanced Analytics: A Point-of-View

Predictive Analytics and Data Mining refer to the use of non-trivial, quantitative means of insight extraction from vast amount of data for the purpose of making informed decisions.

It typically involves data and a quantitative model and often leverages intensive computing and automation.

Some Notable Characteristics

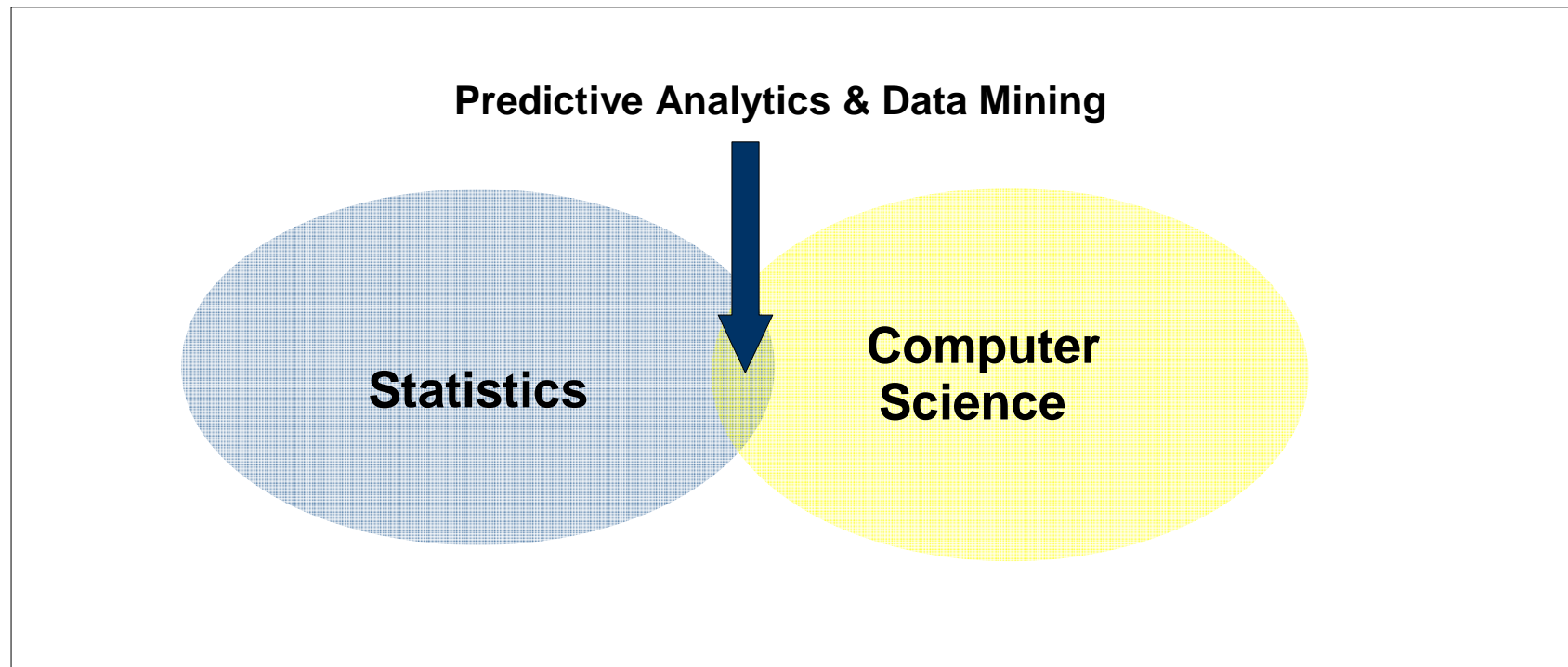
- Non-triviality: if it is easy by contemporary means then it is not advanced (e.g. simple SQL based reporting is not Advanced Analytics)
- Quantitative means: use of scientific techniques; a Mathematical model
- Analysis domain vastness: often driven by data deluge
- Leverage of automation
- Aid of (not yet an alternative to) human decision making

Categories of Advanced Analytics Techniques

- **Predictive Analytics** – predict the future based on existing data points (i.e. past and present)
- **Optimization** – find the best (optimized) approximate solution for a constraint-bound problem
- **Simulation** – explore and understand the behavior of a complex system by studying a mini-replica that mimics reality
- **Visualization (or other forms of Man-Machine interface)** – explore and understand complex systems and its outputs via visual interfaces (often interactive and system-aided)
- **Business Rules** – automated discovery and dynamic implementation of business rules

Predictive Analytics – An Inter-disciplinary Approach

- Based on the past and present, how can we predict the future?
- How can we discover the unknown patterns in my existing data?



Predictive Analytics & Data Mining – Core Concepts

- Classification
- Clustering and Segmentation
- Association / Rule-based
- Forecasting



Next, let us look at each one of concept with some common examples.

What is Classification?

Systematic arrangement in groups or categories according to established criteria

- Whom does it look like? Find the closest match for an item to some pre-defined classes
- Once the match is found, we deal with the item as “one of a kind”



Input item



Is the input part
of this class?



Pre-defined class = Clydesdale



Leverages supervised learning - assumes that prior knowledge of the classes already exist

What is Clustering / Segmentation?

- Assignment of a set of observations into subsets (called clusters) that exhibit similar characteristics
- Divides and Conquers



Input item = People traveling



Is the input part
of this class?



Segments assigned due to destination



Leverages unsupervised learning - assumes that no prior knowledge of the segments exist

What is association?

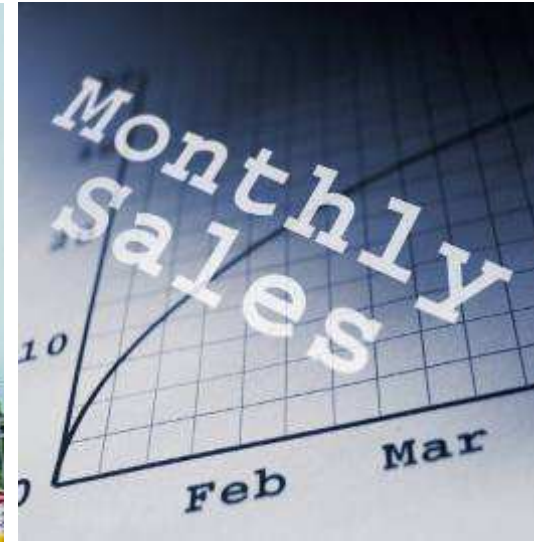
Connects the dots; associate certain data patterns with others based on the similarity between patterns



Friday – 3pm



Friday – 8pm



Monthly Beverage Sales

Is there an association between these items?

Predictive Analytics – Categories

Categories	Purpose	Example
Classification	<p>Whom does it look like? Find the closest match for an item to some pre-defined classes; once the match is found, we deal with it as “one of a kind”</p> <p>Remarks: supervised learning; prior knowledge of the classes exist</p>	<p>I am a Credit Card Approver and I need to classify the new Credit Card Applicant into one of the following classes – Cash Cow, Probable Defaulter or No-profit types. If the applicant is probably more like a Cash Cow type – I shall approve the applicant.</p>
Clustering (Segmentation)	<p>Divide and conquer; break up a large glob of items into some segments or clusters based on the mutual similarity between the items;</p> <p>Remarks: unsupervised learning; no prior idea of segments</p>	<p>I do not even know that what all types or classes exist in my credit card customer database. I want clustering algorithms to find them out based on mutual comparison of the customer behavior profiles</p>
Association	<p>Connect the dots; Associate certain data patterns with the other based on the similarity between patterns;</p>	<p>Friday evening, middle-aged dads buying beer along with diapers on their way out of grocery.</p> <p>Associating the diaper buying behavior with that for beer, for certain demographic, temporal and house-holding profile.</p>
Forecasting	<p>Tell the future; look at the trends in the past and present and predict the future</p>	<p>Forecast the spending pattern of my credit card customer this Christmas based on his income, payment and spend patterns for last 5 years.</p>

Data Mining

Classification per Nature of Data

May be classified as

- Structured (e.g. in-house RDBMS, organized file system)
- Unstructured (e.g. email, public domain Internet data – wikis, blogs, message boards)

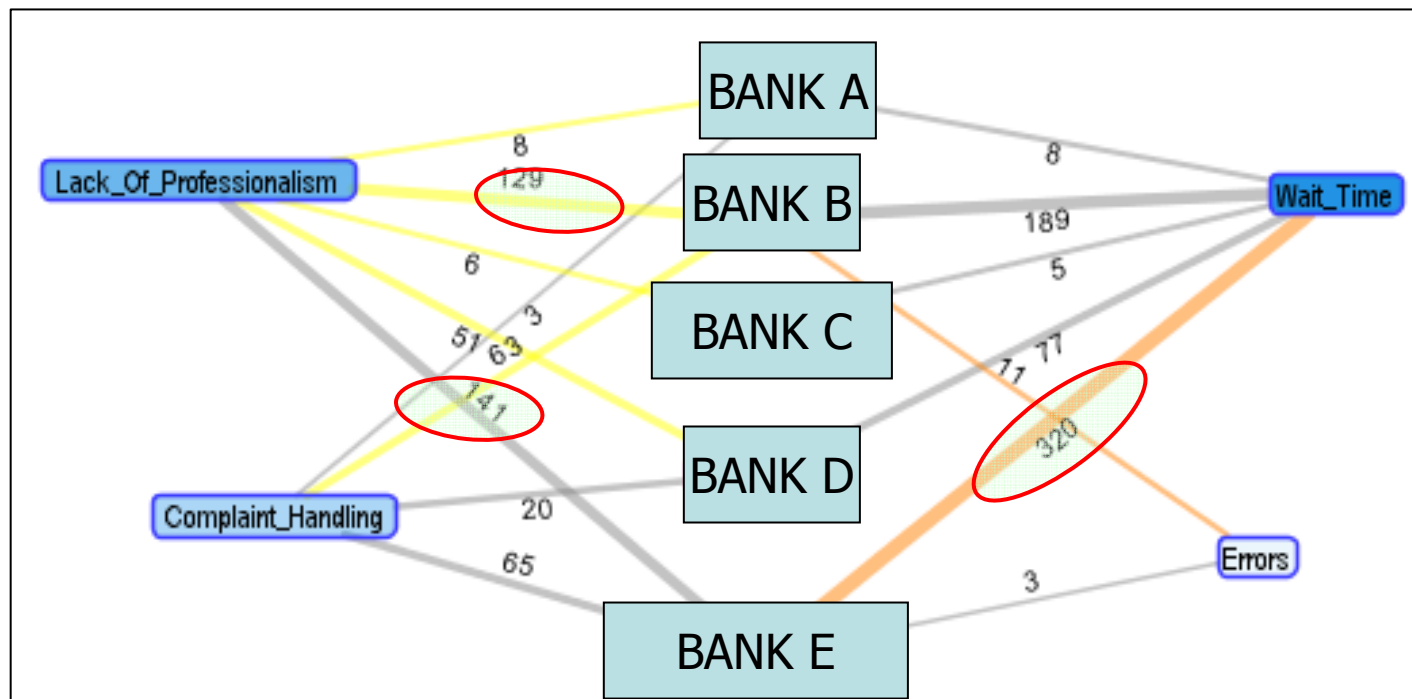
*Let us look at what we can do with **Text Mining** – a specific type of Data Mining – on unstructured data that is publicly available on the World Wide Web.*

Unstructured Data Mining: A Sample Study

- Investigation targets (Banks)
 - Bank A
 - Bank B
 - Bank C
 - Bank D
 - Bank E
- Investigation Topics
 - Wait Time
 - Complaint Handling
 - Lack of Professionalism
 - General Errors
- Sentiments
 - Positive
 - Negative
 - Neutral

Unstructured Data Mining: Public Sentiment Analysis for some banks using IBM COBRA (Corporate Brand and Reputation Analysis Software)

Affinity diagram: relationship between banks and public perception of their services



Some Examples



The Operational Risk Data eXchange

A consortium of 50 leading financial institutions improves statistical modeling and more accurately quantifies risk exposure



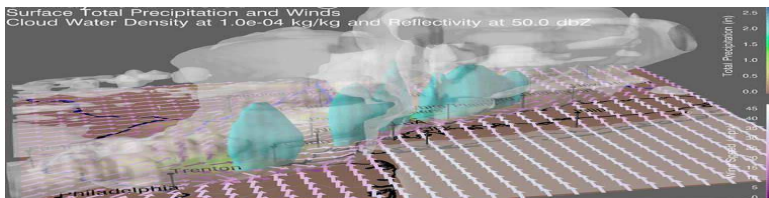
Smarter Policing

NYPD Precinct Commanders are able to take proactive measures and stay ahead of trends and respond to spikes in criminal activity.



Smarter traffic

Singapore - the ability to predict where traffic jams will occur at any given hour minimizes congestion and reduces carbon levels



Deep Thunder

The ability to accurately predict weather for a specific location with business impact analysis