

An Oracle White Paper
September 2010

Predictive Analytics: Bringing The Tools To The Data

Introduction	2
What Are Predictive Analytics -- The Traditional View.....	3
A Better Classification.....	4
Predicting the Present.....	5
Shaping The Future	7
A Better Approach.....	9
Words of Warning	11

Introduction

"It's hard to make predictions, especially when they are about the future" is a quote usually attributed to American baseball-legend Yogi Berra¹.

Not a good start when discussing *predictive analytics*.

It gets even more problematic when we realize what analytics exactly are. Analytics are the methods of decomposing concepts or substances into smaller pieces, to understand their workings. How can you analyze the future? There is no concept or substance to break down yet.

To go further, predictive analytics even sounds like an oxymoron. *Predictive* is rehearsing an important conversation beforehand: "If they say this, I will respond with that...", while *analytics* is evaluating the conversation afterwards; "Gosh, I should have said that..."

It is easy to come to conclusions like this, when you don't have a good understanding of what predictive analytics actually mean. In this paper we discuss what predictive analytics can and cannot do, the various categories of predictive analytics, and how to best apply them.

¹ The quote has been attributed to the American author Mark Twain and the Danish physicist Niels Bohr as well.

What Are Predictive Analytics -- The Traditional View

"Predictive analytics encompasses a variety of techniques from statistics, data mining and game theory that analyze current and historical facts to make predictions about future events"², is how Wikipedia describes it. The variety of techniques is usually divided in three categories: *predictive models*, *descriptive models* and *decision models*. Predictive models look for certain relationships and patterns that usually lead to a certain behavior, point to fraud, predict system failures, assess credit worthiness, and so forth. By determining the explanatory variables, you can predict outcomes in the dependent variables. Descriptive models aim at creating segmentations, most often used to classify customers based on for instance sociodemographic characteristics, life cycle, profitability, product preferences and so forth. Where predictive models focus on a specific event or behavior, descriptive models identify as many different relationships as possible. Lastly, there are decision models that use optimization techniques to predict results of decisions. This branch of predictive analytics leans particularly heavily on operations research, including areas such as resource optimization, route planning and so forth. Table 1 provides an overview.

PREDICTIVE MODELS	DESCRIPTIVE MODELS	DECISION MODELS
<ul style="list-style-type: none"> • Find causality, relationships and patterns between explanatory variables and dependent variables • Focus on specific variables • Examples: next customer preference, fraud, credit worthiness, system failure 	<ul style="list-style-type: none"> • Find clusters of data elements with similar characteristics • Focus on as many variables as possible • Examples: customer segmentation based on sociodemographic characteristics, life cycle, profitability, product preferences 	<ul style="list-style-type: none"> • Find optimal and most certain outcome for a specific decision • Focus on a specific decision • Examples: critical path, network planning, scheduling, resource optimization, simulation, stochastic modeling

Table 1: Practical classification

This classification is very practical; it provides an immediate understanding of the areas where predictive analytics add value. However, there are two problems with it:

- The classification is not exhaustive. If a new area of predictive analytics were to arise tomorrow the current classification would become invalid
- The classification doesn't tell what the categories *don't* do. The limitations of each category are not clear.

² www.wikipedia.com, "Predictive Analytics", August 2010

A Better Classification

The reason why there is so much misunderstanding about predictive analytics is in the common conception that predictions have to be about the future. A more conceptual classification solves this misconception and addresses the issues with the current segmentation. This conceptual classification distinguishes between two types of predictive analytics:

- Predicting the Present
- Shaping the Future

Trends develop like S-curves. They start slowly, take off, and become the new paradigm and "best practice". Then, at one moment, something unexpected happens³. New regulations, a better technology, a scandal within the market, whatever. The list of possible disruptions is endless. Within the paradigm we can predict what is happening, as long as the assumptions don't change. Once there is a disruption, a structural break, a new S-curve builds, and the process repeats. See figure 1.

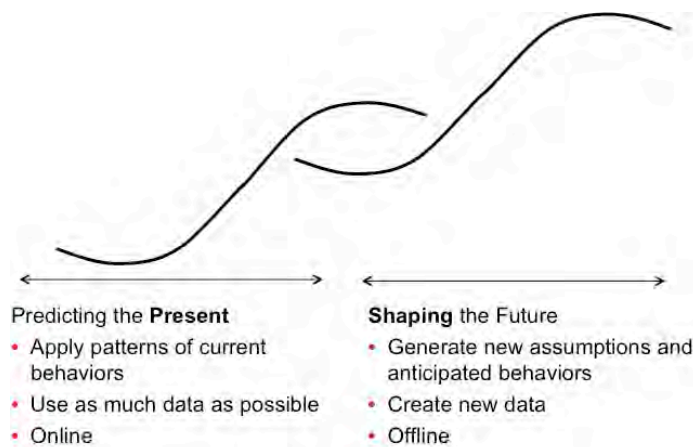


Figure 1: S-curves and predictive analytics.

Predictive analytics that predict the present focus on analyzing what happens within a stable and current situation, based on a set of assumptions that describe reality. Predictive analytics that help shape the future focus on preparing for the *next* S-curve, and help formulate a new set of assumptions.

³ This moment is often referred to as a "Black Swan". Something we thought couldn't happen within the rules of the game.

Predicting the Present

Predictive analytics that predict the present are based on existing data, preferably as much as possible. Through data mining and all other applicable techniques, patterns are detected and rules are derived. Within the bounds of its present S-curve, there is predictive value. Not towards the future, but towards *similar occurrences*. For instance, if Customer A buys a pair of trousers, there is a large chance he or she is also interested in socks. Or if there are expense items with the name of a fashion brand on a corporate credit card, it is an irregular transaction. The analytics take place as deep in the process and as real-time as possible. Table 2 provides an overview of typical uses, based on the functionality of Oracle Data Mining, which is embedded in the Oracle database.

FUNCTIONALITY	ALGORITHM	APPLICABILITY
Classification	Logistic Regression (GLM) Decision Trees Naïve Bayes Support Vector Machine	Response modeling Recommending "next likely product" Employee retention Credit default modelling
Regression	Multiple Regression (GLM) Support Vector Machine	Credit scoring Customer profitability modeling
Anomaly detection	One Class SVM (Support Vector Machine)	Claims fraud Network intrusion
Attribute importance	Minimum Description Length (MDL)	Surgery preparation, triage Net promoter score
Association rules	Apriori	Market basket analysis Link analysis
Clustering	Hierarchical K-Means Hierarchical O-Cluser	Customer segmentation Gene and protein analysis
Feature extraction	Non-Negative Matrix Factorization (NMF)	Text analysis, search

Table 2: Functionality overview Oracle Data Mining

Oracle Data Mining

Oracle Data Mining uses a multitude of algorithms that can automatically sift deep into your data at the individual record level to discover patterns, relationships, factors, clusters, associations, profiles, and predictions—that were previously “hidden”. Since Oracle Data Mining functions reside natively in the Oracle Database kernel, they deliver high performance, high scalability and security. The data and data mining functions never leave the database to deliver a comprehensive in-database processing solution. With Oracle Data Mining, special departments of advanced data analysts working in silos far away from the database are no longer needed.

The true value of data mining is best realized when the new insights and predictions are directly supplied to existing business applications. With Oracle Data Mining users can supply predictive analytics to business applications, call centers, web sites, campaign management systems, automatic teller machines (ATMs), enterprise resource management (ERM), and other operational and business planning applications. The database becomes more than just a data repository—it becomes an *analytical* database that can undergird many new advanced use cases.

Overview of benefits:

- Mines data inside Oracle Database, an industry leader in performance and reliability
- Eliminates data extraction and movement
- Provides a platform for analytics-driven database applications
- Provides increased security by leveraging database security options
- Delivers lowest total cost of ownership (TCO) compared to traditional data mining vendors
- Leverages 30+ years of experience of ever advancing Oracle Database technology

Oracle Real-Time Decisions

Where Oracle Data Mining integrates in the database, as a decision framework Oracle Real-Time Decisions (Oracle RTD) nests itself in process engines. Oracle RTD, at its core, is a closed-loop recommendation engine. It helps optimize customer interactions using analytical techniques such as business rules, data mining, or statistics. Instead of deep analysis in an offline environment -- the database -- Oracle RTD immediately influences the real-time data stream by providing recommendations on how to complete the transaction.

Oracle RTD particularly focuses on personalized customer interactions flows. By learning from every single interaction and adjusting the business processes in real-time, Oracle RTD optimizes the value of each opportunity in itself, as well as every subsequent opportunity. It starts with a list of possible next actions, and then adds information about the customer, session history, customer history, your business goals, and what has worked or not worked in the past. RTD uses that information to come up with a recommendation for the best action to take next. Each decision is made using a *statistical model* that predicts the outcome of the interaction based on what is known about the customer and how similar customers have responded in the past. Because RTD operates as a business process engine, its decision framework can leverage the real-time context of the interaction in rules or predictive models. This allows to account for data such as the time of the day, the agent you are interacting with or the context of your web session to improve the recommendation logic. Context data is a key predictor.

Oracle RTD applications can be fully automated, requiring no manual effort for building and maintaining its predictive models. Oracle RTD automatically learns from each customer interaction by autonomously updating its predictive models in real time.

Shaping The Future

Predicting the present helps an organization improve its organizational excellence. In other words, it increases efficiency and improves the effectiveness of operations. However, organizations also have another set of processes: management processes. These are aimed at creating and implementing new strategies, and monitoring progress. Shaping the future improves an organization's *management excellence*⁴.

One of the most dominant predictive tools on the strategic levels is a strategy map, part of the balanced scorecard. Strategy maps aim to be predictive, as they aspire to show how decisions made in the present could impact future results. This is done through linking leading and lagging indicators. A leading indicator predicts future performance; a lagging indicator reports past performance. For instance, for a postal service, the percentage of mail delivered within 24 hours is a leading indicator for customer satisfaction. Statistics are used to discover and test these relationships. You can use statistical techniques only if you have enough data. Data, by definition, describes results from the past. Given that all we can truly predict about the future is that most likely it will be different from today, you can question the predictive value of correlations found in data describing the past. Putting it in stronger words, you could even argue that validating a strategy map based on correlating past data, by definition, *invalidates* it⁵.

A different approach than predicting the present is needed; an approach in which we *create* the data for the future, in terms of new assumptions, what-if questions and scenarios. It is within these assumptions that the risk and uncertainty that are part of any strategic planning can be captured. To create good assumptions, organizations must complete three critical steps⁶:

- Make the assumptions realistic.
- Identify the assumptions that matter most.
- Leverage the drivers that can be controlled and monitor those that cannot.

Several technologies can be helpful in this process.

⁴ See all of Oracle's management excellence white papers on www.oracle.com/thoughtleadership, in the resource center

⁵ Based on Buytendijk, F.A., "Dealing with Dilemmas", Wiley, 2010

⁶ Driving Strategic Planning with Predictive Modeling, July 2008, www.oracle.com

OLAP

Hyperion Essbase is an OLAP database, in which data is stored in a multidimensional format. This means the data is categorized in the form of dimensions, representing business structures. Typical business dimensions are 'product', 'customer', 'region' and 'time'. To perform what-if analysis and model possible future scenarios, a 'scenario' dimension is added. Scenario-building consists of using the following critical functionalities:

- Create multiple hierarchies. By modeling alternative ways of segmenting customers, grouping products, consolidating organizational entities or any other type of business hierarchy, results can be tested before they are implemented as the new standard way of looking at the business
- Use forecast calculations. Based on multiple assumptions, forecasts can be created. Based on likelihood and realism the best forecast can be determined and implemented.
- Data entry. There is no data about the future available in business systems. Based on new assumptions business users may have to input the data themselves, preferably in a collaborative manner. Multi-user read write capabilities and version control are needed to manage such a process.

Hyperion Strategic Planning

Financial modeling is a special activity. It requires specific functionality. At the same time the structure in which the modeling takes place is highly standardized: a balance sheet, a profit and loss statement and a cash flow overview. Oracle Hyperion Strategic Finance integrates strategic planning into an enterprise planning process. It allows users to quickly develop financial models, perform on the fly what-if impact analysis based on dynamic decision variables and arrive at targets that can then be implemented. The what-if analysis toolkit is a unique and powerful set of out-of-the-box tools that allow you to easily create an unlimited number of scenarios by business unit. Other capabilities let you evaluate any metric's sensitivity to key performance drivers and run periodic "goal-seek" checks to determine the performance level needed to achieve specific financial objectives.

Hyperion Strategic Planning allows organizations to spend more time simulating long-term alternative strategies, developing contingent scenarios, and stress testing financial models.

Crystal Ball

Shaping the future is done outside operational business systems. It requires a modeling framework that can handle both structure and creativity. Crystal Ball uses a spreadsheet interface – the most widely used business analyst tool – to create the data for the future, in terms of new assumptions, what-if questions and scenarios. Either standalone in a spreadsheet, or integrated with Essbase and Strategic Planning, Crystal Ball facilitates the creation of good assumptions.

Going through the three critical steps that are needed to shape the future, Crystal Ball can:

- **Make the assumptions realistic.** Regardless of how much data you have (or none at all) apply any or all of the available techniques to improve the model assumptions: time-series forecasting, regression analysis, distribution fitting and simulation methods.
- **Identify the assumptions that matter most.** Determine the sensitivity of the forecast to each assumption, so you know where to focus your resources. Sensitivity charts and Tornado charts rank and help visualize which assumptions are the most important to the least important in the model.
- **Leverage the drivers that can be controlled and monitor those that cannot.** Because of the uncertainty inherent in models that shape the future, changing a driver (a decision variable) can have a significant effect on the forecast results. For one or two drivers, use the Decision Table tool, which quickly runs multiple simulations to test the effects of changing the drivers. For models that contain more than a handful of decision variables, or where you are trying to optimize the forecast results, optimization is used to automatically find optimal solutions to simulation models.

A Better Approach

Traditionally, predictive analytics require a laborious process. Figure X shows a typical flow of activities.

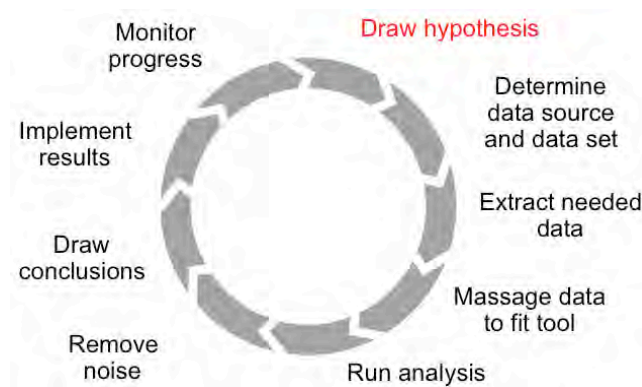


Figure 2: Traditional predictive analytics process

Based on a hypothesis -- an assumption to test -- the data that are needed are identified and extracted from their sources, such as an ERP or CRM system, or coming from the data warehouse. Different predictive analytics tools have different requirements on how to best process the data, usually the data require some transformation to fit the specifics of the tool. Only then the analysis can take place successfully. After the analysis and removal of all the "noise" in the data, conclusions are drawn that lead to changes in for instance customer segmentation or product clustering, and the success of the analysis is monitored. After that, the cycle starts all over.

The key assumption of this approach is that it is best to take the data to the tools, which are handled by the experts. This assumption is outdated. It is a costly approach, takes too much time, robs the process of its creativity and doesn't allow the number of experiments to scale.

Oracle's strategy is different. Oracle feels it is better to bring the tools to the data. Oracle Data Mining resides in the Oracle database itself, and is used by many business applications. Real-Time Decisions nests itself into the process engine of business applications, generating real-time alerts, particularly in customer contact interactions. Hyperion Strategic Finance is an add-on to Hyperion Financial Management and Hyperion Planning, for long-term financial modeling, and Crystal Ball integrates with Essbase and Hyperion planning. Figure 3 shows how the tools are brought to the data.

There are multiple advantages to bringing the tools to the data:

- It is less costly and increases speed. There is no need to extract and transform data and move it to another environment, which is often the most laborious part of the process.
- Moreover, the increased speed of the process makes the process more iterative, and therewith increases the quality of the output. It becomes possible to run many more analyses and scenarios in the same amount of limited time, before deciding what the best outcome is.
- Predictive analytics become part of a standard business process, instead of a specialist activity, and has a more direct business impact.
- It adds creativity to the process. If attributes need to be added while experimenting with multiple analyses, the addition can be done without going back to the source data. The hypothesis can be narrowed and widened without significant consequences.

In addition, Oracle Real-Time Decisions, Crystal Ball, Hyperion Essbase and Strategic Finance integrate with non-Oracle databases and non-Oracle business applications as well.

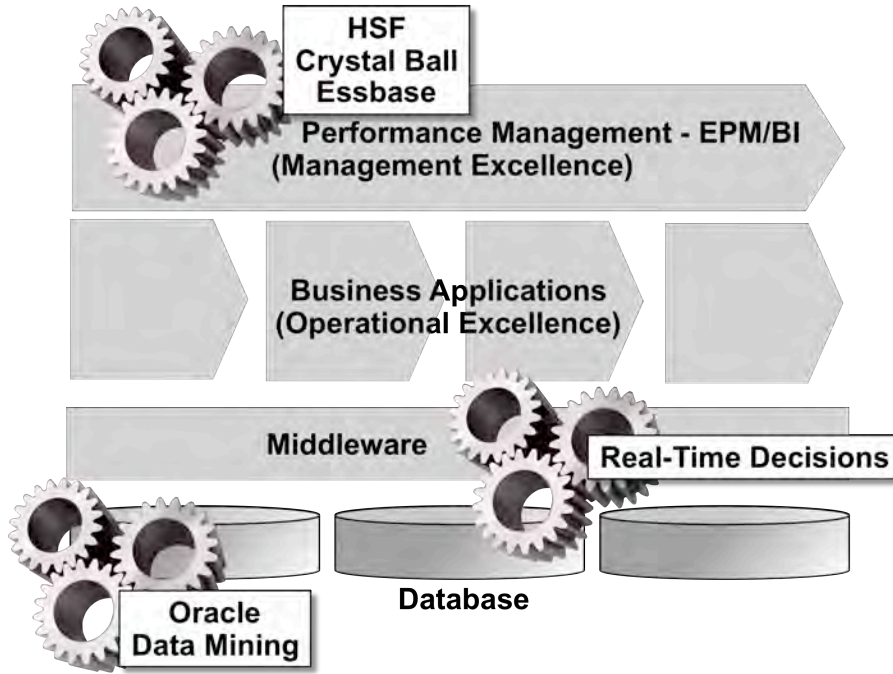


Figure 3: Bringing the tools to the data

Words of Warning

To truly understand the possibilities and limitations of such advanced technologies on which predictive analytics are based, ironically we need to go back to the old philosophers. Plato's Cave describes an interesting thought experiment. Imagine a few people in a cave, sitting against a small wall, facing the other end. They are locked in chains and have been for all their lives. On top of the wall other people walk around, but they cannot be seen by the chained people. A large fire in the back of the cave casts the shadows of those people walking around on the wall at the other end, and the echos of their voices makes the sound come from the other end too. For the people in chains these shadows are the reality of the world, they do not have any knowledge about the people walking around on top of the little wall. The shadows are all they know.

Analysts can easily fall into the same trap, unconsciously mistaking their model and analysis for the real world. Reality is interpreted "to fit the model", signals that change is coming are called

"outliers". Force-fitting reality into our frame of reference is deeply human, and we are all prone to it. What helps counter this pitfall is to experiment⁷. Continuously try the outcomes of predictive analytics in your operational environment. Even run multiple experiments at the same time. See which ones lead to better results, and implement. Using simulations and other techniques, predictive analytics should be applied continuously and incrementally.

Another word of warning comes from "Occam's Razor", based on the work of the 14th century Franciscan friar William of Ockham. Occam's razor is a research principle. In forming a theory, you should use the fewest elements needed to accurately predict an outcome. The more assumptions you use, and the more you stack analysis on top of analysis, the more likely you are to be inaccurate or wrong. To create a robust theory, you need to shave away all unnecessary elements from the analysis (hence "razor"). Seen this way, two of the least reliable indicators of business success are profit and customer satisfaction. They are based on many assumptions. Better indicators are for instance cash-flow and on-time delivery, that are much closer to direct observation and measurement.

Although predictive analytics are based on sophisticated technologies and mathematical techniques, their success lies in applying them with common sense.

⁷ "Analytical competitors", as they are called in Davenport's book "Competing on Analytics", take this approach to analytics.



Predictive Analytics: Bringing The Tools To The
Data

September, 2010

Authors: Frank Buytendijk, Lucie Trepanier

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:

Phone: +1.650.506.7000

Fax: +1.650.506.7200

oracle.com



Oracle is committed to developing practices and products that help protect the environment

Copyright © 2010, Oracle and/or its affiliates. All rights reserved. Published in the U.S.A. This document is provided for information purposes only, and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

10036922