Gildan Media

Companion PDF

PREDICTIVE ANALYTICS

by

Eric Siegel

See also this book's website for all the book's figures,
extensive excerpts (opening chapters and more),
related articles, videos, and more:
www.thepredictionbook.com.

182 Examples of Predictive Analytics

A Cross-Industry Compendium of Mini-Case Studies

For citations pertaining to each of these examples, see the Central Tables Notes at www.PredictiveNotes.com.

Table 1 Predictive Analytics in *Family and Personal Life*

What's predicted:	Example organizations that use predictive analytics:
Location *where you will be*	**Nokia:** Sponsored a competition to predict future location from cell phone tracking. Resulting methods predict location one day beforehand within 20 meters, on average, among a population based in a certain region of Switzerland, in part by incorporating the behavior of friends (i.e., the social contacts whom one calls). **Microsoft:** Helped develop technology that, based on GPS data, accurately predicts one's location up to multiple years beforehand.
where you are headed	**Uber:** Can predict the specific destination address of San Francisco riders based on exact drop-off location with 74 percent accuracy, despite, for example, how many businesses there are within 100 meters in a typical city area (just taking the closest candidate address achieves 44 percent accuracy).
Who's in a photo (aka *facial recognition*—𝒟)	**Facebook:** Improved the state of the art for identifying people from photos to virtually the same performance level as a human: Given two face images, it can determine whether they're the same person with 97 percent accuracy. Facial recognition helps users tag photos, which they do more than 100 million times a day. The company has also developed predictive models to identify people even if it can't see the face, achieving 83 percent accuracy when faces are at least partially obscured half of the time, based on elements such as clothing, hair, and pose.
Which Facebook posts you will like *in order to optimize your news feed*	**Facebook:** Predicts which of 1,500 candidate posts (on average) will be most interesting to you in order to personalize your news feed. To optimize the order of content items, the *News Feed ranking algorithm* weights around 100,000 factors such as recency, likes, clicks, shares, comments, time spent on posts, poster popularity, your affinity for the poster and content area, and measures of relevance and trustworthiness. This intensifies the "addictive" engagement, with two-thirds of Facebook's 1.44 billion monthly users logging in daily.
Acceptance of booking request *in order to match guests to hosts*	**Airbnb:** Rank orders accommodations that fulfill a user search in part by the predicted probability each host would accept the user's booking request. By surfacing likely matches more prominently, the company increased booking conversions by nearly 4 percent—a significant gain considering its estimated annual booking of over 12 million guest nights. *For another Airbnb example, see* Table 3.
Friendship (𝒟)	**Facebook:** Sponsored a competition to improve the precision of suggested people you may know and wish to link to. **LinkedIn:** Considers its predictive suggestions of people you may know "the most important data product we built."

What's predicted:	Example organizations that use predictive analytics:
Love	**Match.com:** For "Intelligent Matching" in online dating, predicts which prospective matches you'd be interested in communicating with. **OkCupid:** Predicts which online dating message content is most likely to elicit a response.
Pregnancy (𝒟)	**Target:** Predicts customer pregnancy from shopping behavior, thus identifying 30 percent more prospects to contact with offers related to the needs of a newborn's parents (*more details in* Chapter 2).
Infidelity	**University researchers:** Showed that cheating in a relationship is predicted more by behavioral traits than by demographic profiles, but that genetic factors are also in play.
Divorce	**Clinical researchers:** Predict divorce with varying degrees of accuracy.
Death	**See tables that follow for examples in *Insurance, Healthcare, Law Enforcement, and Safety.***

𝒟 Rather than performing *prediction* in the conventional sense of the word, this application of predictive modeling performs *detection*. As with predicting the future, such applications imperfectly infer an unknown—but in this case, the unknown could be already known by some, rather than becoming known only when witnessed in the future. This includes predicting the answer to a question, predicting a diagnosis, predicting whether a transaction is fraudulent, and predicting what a subject is thinking.

Table 2 Predictive Analytics in Marketing, Advertising, and the Web

What's predicted:	Example organizations that use predictive analytics:
Purchases *in order to target marketing* (aka *response modeling*)	**PREMIER Bankcard:** Reduced mailing costs by $12 million. **First Tennessee Bank:** Lowered mailing costs by 20 percent and increased responses by 3.1 percent for a 600 percent return on PA investment. **Target:** Increased revenue 15 to 30 percent with predictive models. **Harbor Sweets:** Analytically targeted lapsed customers to win them back at an impressive 40 percent response rate. **Fingerhut:** Targeting reduced direct mailing 20 percent, saving almost $3 million annually, yet increasing earnings. **Vermont Country Store:** More precisely targeting catalog mailings earned a return 11 times higher than the investment to do so. **Harrah's Las Vegas:** This casino predicts how much a customer will spend over the long term (aka *lifetime value*). **Cox Communications:** More than tripled direct-mail response rate by predicting propensity to buy. Predicting demand for communications products such as home TV, Internet, and phone services achieves a 50 percent annual return. **A mutual-fund investment management firm:** Identified clients five times more likely than average to make additional investments. **U.K. supermarket:** Can predict the exact date customers will return and the amount they will spend within US$10, for 19 percent of customers. **Elie Tahari:** Forecasts demand for women's fashion line products. **Life Line Screening:** Increased response 38 percent and cut campaign costs 28 percent for direct mail offering health screenings, a service the firm provides over 1 million times per year.
Cancellations *in order to retain customers* (aka *churn modeling*)	**PREMIER Bankcard:** Retained $8 million in customer value. **FedEx:** Predicts which customers will defect to a competitor with 65 to 90 percent accuracy. **Optus (Australia):** Identified cell phone subscribers 10 times more likely than average to cancel. **Sprint:** Identified telecom customers three times more likely than average to cancel.

What's predicted:	Example organizations that use predictive analytics:
	Telenor (Norway): Reduced cell phone subscriber turnover 36 percent; retention process return increased elevenfold (*more details in* Chapter 7). **2degrees (New Zealand):** Identified cell phone subscribers over 12 times more likely than average to cancel. **Lloyds TSB:** Increased annual profit £8 million by improving the predictive modeling of customer defection. **Chase:** *See entry in* Table 3*, Financial Risk and Insurance* **Reed Elsevier:** Gained 16 percentage points in renewal rate for a magazine. **PayPal:** *See Table 8*
Successful sales *in order to prioritize sales leads*	**IBM:** IBM Canada predicts whether planned sales-oriented events will meet attendance goals with 83 percent confidence—"*If we host the party, will enough people show up?*" This includes IBM's sale of PA capabilities, so this effort has PA selling itself. **Hewlett-Packard:** An early warning system alerting sales staff of business account opportunities predicts outcomes for 92 percent of sales efforts with an accuracy of 95 percent, and predicts the timing of sales closures 60 percent of the time. **Bella Pictures:** Targets brides-to-be for photography services. **Paychex:** This payroll processor decreased by 40 percent the number of phone calls needed in order to book each sales meeting, thus increasing overall sales. **Sun Microsystems:** More than doubled the number of leads per phone contact.
Product choices *for personalized recommendations*	**Amazon.com:** Thirty-five percent of sales come from product recommendations. The company may also develop "anticipatory shipping" that would proactively place packages before they are ordered at hubs or on trucks in order to reduce delays between ordering and receiving purchases, for which it has obtained a patent. **Netflix:** Sponsored a $1 million competition to improve movie recommendations; a reported 70 percent of Netflix movie choices arise from its online recommendations (*more details in* Chapter 5). **Tesco (U.K.):** Annually issues 100 million personalized coupons at grocery cash registers across 13 countries. Predictive modeling increased redemption rates by 3.6 times, compared to previous methods. **Target:** Increased revenue 15 to 20 percent by targeting direct mail with product choice models.

What's predicted:	Example organizations that use predictive analytics:
	U.S. Bank: Response doubled, cross-sell return on investment increased fivefold. **Pandora:** Recommends songs based on 400 musical attributes. **Spotify:** Is augmenting its song recommendation algorithm to incorporate musical attributes.
Clicks *in order to select which ad or content to display*	**Google:** Improves search functionality by predicting which Web pages will meet users' high-quality standards if shown as search results. **Facebook:** In order to increase revenue from its pay-per-click advertisers, predicts ad clicks based on user attributes, device used, and contextual factors. **Education portal:** Increased ad revenue rate by $1 million every 19 months by displaying the ad you're more likely to click (*more details in Chapter 1*).
Ineffective ads *to warn paying advertisers accordingly*	**Google:** Predicts which new ads will get many bounces (when people click on an ad but then immediately click the back button).
Viral tweets and posts *for maximal publicity*	**MTV:** Achieved a 55 percent increase in Web page views when publicizing the Video Music Awards.
Spam *to send it to your spam folder*	**Google:** Decreased Gmail's prevalence and false positive rate of spam from disruptive (in 2004) down to negligible.
Hit songs and movies	**Researchers:** Employ machine learning to predict which screenplays will be Hollywood blockbusters and which songs will hit the charts—*For example, see the Epagogix entry in Table 3, Financial Risk and Insurance.*

Table 3 Predictive Analytics in *Financial Risk and Insurance*

What's predicted:	Example organizations that use predictive analytics:
Bodily harm from car crashes	**Allstate:** With a predictive modeling competition in 2012, tripled the accuracy of predicting bodily injury liability based solely on the characteristics of the insured vehicle. This could be worth an estimated $40 million annually to the company.
Costly workplace injuries	**Accident Fund Insurance:** Ascertains secondary medical conditions (such as obesity and diabetes) from written workers' compensation claim notes. These conditions are predictive of which injuries will be high-cost so that, for example, insured workers may be targeted for preventive measures.
Insurance claims	**Infinity Insurance:** *See the entry in Table 7 for "Application approvals and denials."* **Leading international commercial lines insurance provider:** Predictive models decreased the loss ratio by a half point, contributing to savings of almost $50 million.
Death	**Life insurance companies:** Predict age of death in order to decide upon policy application approvals and pricing. **A top-five U.S. health insurance company:** *Death prediction is not within the usual domain for health insurance—see the Healthcare table below for the nature of this work.*
Mortgage prepays	**Chase:** Generated millions of dollars with predictive models that foresee which homeowners will refinance their mortgages and thereby take all future interest payments to a competing bank (*more details in Chapter 4*).
Loan defaults (risk)	**Citigroup:** Leverages over 30 years of international loan default history to generate commercial credit-risk models for individual regions (for North America and Western Europe, it breaks down further into industry-specific models), for which there are up to 3,000 internal users; models have been in existence for over 20 years. **Canadian Tire:** Predicts late credit card bill payments in order to manage risk. **PREMIER Bankcard:** Lowered delinquency and charge-off rates, increasing net by $10+ million. **Mimoni (Mexico):** Predicts first-loan defaults, thereby reducing first-loan losses by more than 30 percent in comparison to a randomized control group. This in turn enables the organization to fulfill its mission to offer credit to unbanked, working people in

What's predicted:	Example organizations that use predictive analytics:
	Mexico who would otherwise be left only with unfavorable or illegal alternatives. **Kiva:** Detects microloan projects (to alleviate poverty) almost twice as likely as average to default on payback, in part by the absence or presence of key words such as *school* and *machine* in project descriptions. The insights could guide individual lenders, who may lend through www.kiva.org as little as $25.
Nonpayment	**Brasil Telecom (now Oi, which means "hi"):** Predicted bad debt to recover US$4 million. **DTE Energy:** Seven hundred percent increase in net savings (e.g., by preempting charge-offs and decreasing service disconnects). **Financial institution:** Saved $2.1 million in losses by offering collection deals to those accounts that will otherwise not pay, and *not* offering to those that will.
The stock market *for black-box trading*	**John Elder:** Invested all his personal assets into a black-box trading system of his own design (*more details in Chapter 1*). **London Stock Exchange:** An estimated 40 percent of London Stock Exchange trading is driven by algorithmic systems. **Various firms:** Cerebellum Capital, Rebellion Research, and many other firms trade algorithmically.
Airfares	**Hopper:** Predicts airfare changes in order to recommend to consumers whether to buy or wait. Ninety-five percent of these predictions save the consumer money or do no worse than the first price seen, saving users an average 10 percent on ticket price.
Accommodation bookings at a given price–*for dynamic pricing*	**Airbnb:** Suggests each day's price for an accommodation listing (the "Price Tips" feature) by way of predicting whether the listing will be booked—predicted demand directly informs optimal pricing. Bookings are predicted by day of the week, seasonality, and local events, as well as characteristics of the listing such as the neighborhood, size, amenities, key words like "beach," number of reviews, and photographs. Hosts who set prices within 5 percent of the suggestions improve their chance of booking by a factor of nearly four. *For another Airbnb example, see Table 1.*
Blockbuster movies	**Epagogix:** Predicts the success a movie will see if the investment is made to produce it in order to inform studio decisions as to whether to greenlight a film. The company reports success predicting box office gross within $10 million in 83 percent of cases based on actors, director, and characteristics of the script such as certain plot elements.

Table 4 Predictive Analytics in *Healthcare*

What's predicted:	Example organizations that use predictive analytics:
Death	**A top-five U.S. health insurance company:** Predicts the likelihood an elderly insurance policyholder will pass away within 18 months in order to trigger end-of-life counseling (e.g., regarding living wills and palliative care) (*a few more details in Chapter 2*). **Riskprediction.org.uk:** Predicts your risk of death in surgery based on aspects of yourself and your condition—for minor, major, or complex operations, or for certain specialized operations such as a restorative proctocolectomy.
Surgical site infections	**University of Iowa Hospitals and Clinics:** Identifies cases greater than four times as likely to develop surgical-site infections. Targeting anti-infection therapy accordingly reduces the cost of each colorectal surgical procedure an average of $1,300 and will provide a projected annual savings of several million dollars once expanded to other forms of surgery.
Influenza	**Google Flu Trends:** Shown to foresee an increase in influenza cases at one hospital 7 to 10 days earlier than the Centers for Disease Control by incorporating online search trends (e.g., related to symptoms).
Breast cancer (𝒟)	**Stanford University:** Derived with predictive modeling an innovative method that diagnoses breast cancer better than human doctors, in part by considering a greater number of factors in a tissue sample.
Sepsis	**Sisters of Mercy Health Systems:** Predicts severe sepsis and septic shock based on patient vital signs observed over time—detected 71 percent of cases with an acceptable false positive rate.
HIV progression	**Researchers:** Improved the accuracy of predicting disease progression from 70 to 78 percent.
Effect of a drug	**Pfizer:** Predicts the probability a patient will respond positively to pharmaceutical treatment within three weeks.
Premature birth	**Brigham Young University and University of Utah:** Correctly predict about 80 percent of premature births (and about 80 percent of full-term births), based on *peptide biomarkers*, as found in a blood exam as early as week 24 of a pregnancy.

What's predicted:	Example organizations that use predictive analytics:
Erectile dysfunction (𝒟)	**Pfizer:** Derived a more effective, simpler, self-administered diagnostic test of five questions.
Hospital admissions	**Heritage Provider Network:** Awarded $500,000 to a team of scientists who won an analytics competition to best predict the number of days a patient will spend in the hospital over the next year. **University of Pittsburgh Medical Center:** Predicts a patient's risk of readmission within 30 days, in order to assist with the decision to release.
Skipped drug doses	**FICO:** Predicts patient compliance to drug prescriptions, identifying groups that will miss, on average, hundreds of days of regimen per year. Nonadherence to medication prescriptions causes an estimated 125,000 premature deaths and over $290 billion in avoidable costs annually in the United States alone.
Clinical-trial recruitment	**GlaxoSmithKline (UK):** Predicts the supply of much-needed participants for the clinical trial of a new drug in order to plan and allocate expensive trial drug supplies. Clinical trials are a major bottleneck and cost in pharmaceutical research and development (R&D).
Billing errors (𝒟)	**MultiCare Health System (four hospitals in Washington state):** Detected errant accounts and claims to realize $2 million in missed charges within one year.
Various health risks	**Medical centers and healthcare providers:** Proactively target marketing solicitations for preventive and early intervention healthcare to individuals with higher health risks. **Blue Cross Blue Shield of Tennessee:** From claims data, predicts which healthcare resources individual members will need.

𝒟 Detection, not prediction—see Table 1 for more information.

Table 5 Predictive Analytics in *Law Enforcement and Fraud Detection*

What's predicted:	Example organizations that use predictive analytics
Fraudulent: (𝒟)	
Tax returns	**Internal Revenue Service:** Predictively ranking tax returns suspected of cheating empowered IRS analysts to find 25 times more tax evasion, without increasing the number of investigations.
	New York State: Stopped 252,000 bogus refund claims in 2014, saving taxpayers $450 million.
Government invoices	**U.S. Department of Defense's Defense Finance and Accounting Service:** Detected 97 percent of known fraud cases in a published study.
Government contracts	**U.S. Postal Service:** Predictively ranks suspected incidents of contract fraud such as collusion or preferential treatment in order to guide investigations.
Workers' comp claims	**U.S. Postal Service:** Predicted which workers' compensation claims and payments are unwarranted, contributing to a savings of $9.5 million achieved by analytical approaches.
Medicaid claims	**New York City Medicaid administrators:** Increase the probability of fraud investigation success by 71 percent.
Medicare claims	**The Centers for Medicare and Medicaid Services (U.S. federal agency):** Analytically screens every Medicare fee-for-service claim for fraud before issuing payment, gaining a return on investment of 500 percent on the project costs to do so.
Automobile insurance claims	**$40+ billion U.S. insurance company:** Predictively ranking suspected claims empowered auditors to find 6.5 times more fraud with the same number of investigations.
	Aviva Insurance (U.K.): Improved the detection of fraudulent auto claims that include a bodily injury component, amounting to a new savings of half a million pounds per month.
Warranty claims	**Hewlett-Packard:** Realized $66 million in savings over five years by detecting fraudulent warranty claims submitted by HP sales and service partners.
Checks	**Citizens Bank:** Predicted which checks are fraudulent well enough to decrease fraud loss by 20 percent.
Murder	**Maryland:** Applies predictive models to detect inmates more at risk to be perpetrators or victims of murder.
	Baltimore: Among those on parole or probation who will kill or be killed, correctly predicts 75 percent (a *recall* of 75 percent).

What's predicted:	Example organizations that use predictive analytics
Civil unrest	**Israel Institute of Technology:** Researchers created models that predict 51 percent of riots with a precision of 91 percent.
Street crime	**Chicago, Los Angeles, Miami, Memphis, Richmond (VA), Santa Cruz (CA), and Vineland (NJ):** Direct police to patrol areas where crime is predicted (*more details in Chapter 2*).
Terrorism	**The National Security Agency:** Obtained software solutions for and core competency in predictive analytics. It's clear that the NSA considers predictive analytics a strategic priority as a means to target investigation activities by automatically discovering previously unknown potential suspects (*more details in Chapter 2*). **U.S. Armed Forces:** Conduct and fund research to analytically predict terrorist attacks and armed opposition group activities based on factors such as relevant arrests, trials, financial support received, and contextual political conditions.
Ducking city regulations (𝒟)	**City of New York:** Targeting investigations increased by five times the discovery of illegal apartments and flipped business licenses, and doubled the detection of stores selling bootlegged cigarettes.
Recidivism (repeat offenses): *To decide on prison sentencing and parole* *To drive rehabilitation assignments*	**Oregon and Pennsylvania:** Judges and parole boards consult predictive models in order to help decide who stays incarcerated, and for how long (*more details in Chapter 2*). **The Florida Department of Juvenile Justice:** Improved accuracy by about 30 percent over their preexisting, actuarially based risk-assessment tool that drives per-juvenile rehabilitation assignment decisions, achieving an accuracy of 82 percent and reducing false positives by over 40 percent.
Whether a murder will be solved	**Chicago Police Department:** Found that characteristics of a homicide and its victim help predict whether the crime will be solvable.
Security level	**Amazon.com:** Predicts the appropriate security access needs of employees.
Hackers and viruses (𝒟)	**Researchers:** Predictively model which online activities are malicious intrusions and attacks, and which are legitimate activities.

𝒟 Detection, not prediction—see Table 1 for more information.

Table 6 Predictive Analytics in *Fault Detection, Safety, and Logistical Efficiency*

What's predicted:	Example organizations that use predictive analytics:
System failure *to preemptively intervene*	**Large commercial satellite company:** Discovered patterns in satellite battery failures in order to better predict which satellites will require maintenance three years out.
Satellites	
Nuclear reactors (𝒟)	**Argonne National Laboratory:** Predictive modeling of nuclear reactor failures (e.g., cracks in cooling pipes).
City power	**Con Edison:** In New York City, predicts failure of energy distribution cables, updating risk levels that are displayed on operators' screens three times an hour.
Manholes	**Con Edison:** Predicts dangerous manhole explosions and fires in New York City, identifying a 2 percent of manholes that have a 5.5 times greater than average risk of an incident.
Train tracks (𝒟)	**BNSF Railway:** Predicts broken train tracks, the leading cause of severe train accidents, generating location-specific service failure predictions with 85 percent accuracy.
Train wheels	**TTX:** Predicts the failure probability for each of hundreds of thousands of railcar wheels in order to forecast overall annual inventory and maintenance need within a 1.5 percent margin.
Office equipment	**Fortune 500 global technology company:** Predicts which components of electronic equipment such as printers and hard drives are most likely in need of replacement in order to preload repair dispatch trucks.
Credit card payment systems (𝒟)	**Leading payments processor:** A project to detect transaction system behavior anomalies so that problems are resolved more quickly achieved a sevenfold return on investment.
Buildings	**Universities in Iran:** Predict the strength of concrete based on how it is composed and mixed.
Company networks	**18 percent of information technology (IT) departments (poll):** Employ predictive analytics to warn of impending IT failures in order to take corrective measures such as preparing for a spike in online activity.
(*aka reliability modeling*)	
Defective items *for assembly line quality control* (𝒟)	**Washing machine manufacturer:** Achieved a fault-detection performance exceeding 87 percent.
Oil flow rate *in order to efficiently tap underground petroleum reserves*	**National Iranian South Oil Company:** Uses a neural network to predict the rate of oil production.
Oil refinery safety incidents	**Shell:** Predicts the number of safety incidents per team of workers at oil refineries, globally. One example discovery: Increased employee engagement predicts fewer incidents; one percentage point increase in team employee engagement is associated with a 4 percent decrease in the number of safety incidents per FTE.

What's predicted:	Example organizations that use predictive analytics:
Maritime incidents	**RightShip:** Predicts dangerous or costly maritime incidents in order to assess vessel risk that informs shipment decisions when selecting between vessels. The 10 percent highest-risk vessels are three times more likely than average to experience an incident in the next 12 months, and are 16 times more likely to incur a casualty than the 10 percent least risky. Risk assessment is based on vessel age, type, carrying capacity, origin, registration, ownership, management, and other factors.
Deliveries—*which addresses will receive a package*	**UPS:** Cut 85 million miles from annual delivery vehicle driving with a semiautomatic optimization system that plans vehicle/package assignments, as well as package placement within the vehicle, based upon each day's analytically predicted delivery destinations. *See also Amazon.com in Table 2*.
Car service passenger destination	**Uber:** *See entry in Table 1, Family and Personal Life*
Customer need *in order to streamline service*	**Canadian Automobile Association:** New process responding to customer calls reduced dispatches 25 percent and cut labor costs when service isn't needed, yet lowered dissatisfaction 45 percent.
Airplane crash fatalities (𝒟)	**Analytics leaders:** Identify aviation incidents five times more likely than average to be fatal, modeling on data from the National Transportation Safety Board.
Flight delays	**Continental Airlines:** Improved aviation delays and airspace performance by predicting them with radar network data, saving tens of millions of dollars. **A large U.S. carrier:** Can accurately predict more than 25 percent of maintenance delays/cancellations by using maintenance logs, aircraft sensors, and other information about each aircraft and flight, which translates to net cost savings of tens of millions of dollars a year via preemptive response if deployed across the airline's entire fleet.
Traffic	**New South Wales, Australia:** Predicts travel time on Sydney, Australia's M4 freeway. Anticipated travel delays will be provided online to locals just as weather forecasts are.
Dropped calls	**Nokia Siemens Networks:** Predict customer outages on a 4G wireless network with 70 percent accuracy in order to improve service availability and continuity.
Fire and lead poisoning	**Cities of New York and Chicago:** *See entries in Table 7*.
Driver inattentiveness	**Ford Motor Co., Averitt, Air Force:** *See entries in Table 8*.

𝒟 Detection, not prediction—see Table 1 for more information.

Table 7 Predictive Analytics in *Government, Politics, Nonprofit, and Education*

What's predicted:	Example organizations that use predictive analytics:
Voter persuasion	**Obama for America 2012 Campaign:** Predicted which voters would be positively influenced by campaign contact (a call, door knock, flier, or TV ad), and which would actually be caused to vote adversely by contact. Employed to drive campaign decisions for millions of swing state voters, the predictive models "showed significant lift" over traditional campaign targeting (*more details in Chapter 7*). **Hillary for America 2016 Campaign:** Given Obama's success with persuasion modeling in 2012, Hillary Clinton's 2016 campaign appears to be planning to employ it as well. Analytics job postings reveal they're going to be "helping the campaign determine which voters to target for persuasion."
Donations	**Charlotte Rescue Mission:** Increased contributions in response to fund-raising campaigns by nearly 50 percent. **The Nature Conservancy:** Discovered how to profit $669,000 by mailing to only the 10 percent of its donor list predicted to be most likely to contribute. **JustGiving:** Credits predictive analytics as central in an expected increase in fund-raising of hundreds of millions of British pounds. **University of Utah's School of Business:** Increased alumni donations 73 percent by predicting response to annual outreach.
Restaurant health code violations *via Yelp reviews*	**City of Boston:** Sponsored a competition that generated the ability to predict whether a restaurant will have more violations than normal with 75 percent accuracy, in part by way of discovering clues within Yelp reviews, in order to target city health department inspections. Similar work for **Seattle** restaurants distinguished severe violators with 82 percent accuracy.
Lead poisoning from paint (𝒟)	**City of Chicago:** Identified 5 percent of homes that are at more than twice the risk for lead poisoning incidents than average based on the age of the house, the history of lead paint exposure at that address, the economic conditions of the neighborhood, and other factors. This serves as an early warning system to proactively flag, as an improvement over the more common reactive steps taken after a positive test for poisoning. The risk scores serve to target homes for inspection and children for testing, and could help people determine safer homes to move to.

What's predicted:	Example organizations that use predictive analytics:
Fire	**City of New York:** Targets the fire inspections of its 330,000 inspectable buildings with a predictive model that assesses risk based on about 60 factors.
Awarding of grants	**University of Melbourne:** Sponsored a predictive modeling competition to predict which applications for research grants will be approved.
Energy consumption	**Energex (Australia):** The country's second-largest utility spatially simulates 20 years of forecasted electricity demand growth in order to direct infrastructure development and target consumers with incentives to decrease energy consumption.
Overpriced property leases (𝒟)	**U.S. Postal Service Office of Inspector General:** Predicted the amount paid over market value for each of their 26,000 leased facilities (e.g., retail unit, plant, warehouse). Targeting facilities in the Northeast Region, USPS auditors projected that 250 of the leases predicted as most overpaid represent a potential savings of $6.6 million by way of renegotiating their next year of lease terms.
Loan defaults *First loans* *Microloans*	**Mimoni:** *See entry in Table 3, Financial Risk and Insurance* **Kiva:** *See entry in Table 3, Financial Risk and Insurance*
Application approvals and denials	**U.S. Social Security Administration:** Sped up its response to disability claims for a significant subset of applicants, from over a month to under an hour. **Infinity Insurance:** 1,100 percent increase in fast-tracking claims.
The need for help (𝒟)	**British Broadcasting Corporation:** Targeted which home TV viewers were most likely in need of technical assistance with the switch-over to digital TV, especially among the older and disabled.
Dropouts	**American Public University System, Arizona State University, Iowa State University, Netherlands' Eindhoven University, Oklahoma State University, and University of Alabama:** Predict which students are at risk of dropping out in order to intervene and assist in the hope of retaining them.
Grades: *So the computer can grade automatically* *So academic assistance can be targeted*	**Hewlett Foundation:** Sponsored the development of automatic grading of student-written essays. The resulting system grades essays as accurately as (i.e., in agreement with) human graders. **University of Phoenix:** Predicts which students risk failing a course in order to target intervention measures such as adviser coaching. **Rio Salado Community College:** Predicts after eight days of class whether students will attain a C or better with 70 percent accuracy, based in part on online behavior, in order to alert professors.

What's predicted:	Example organizations that use predictive analytics:
Knowledge *for personalized education*	***Jeopardy!* winner:** Analytics expert Roger Craig predicted which practice questions he'd get wrong in order to target his many hours of studying for an appearance on this TV quiz show. He attained the highest one-day winning total ever and won the show's 2011 Tournament of Champions (*more details in Chapter 6*). **Facebook, Elsevier, IBM, and Pittsburgh Science of Learning Center:** Sponsored a predictive modeling competition to predict student performance on algebra problems. Predictively tailored instruction by Intelligent Tutoring Systems promises to save an estimated 250 million student hours per year. **Grockit:** This test preparation company predicts which GMAT, SAT, and ACT questions a test taker will get wrong in order to target areas for which an individual needs more study.

𝒟 Detection, not prediction—see Table 1 for more information.

Table 8 Predictive Analytics in *Human Language Understanding, Thought, and Psychology*

What's predicted:	Example organizations that use predictive analytics:
Answers to questions (𝒟)	**IBM:** Developed with predictive modeling the Watson question-answering computer, which defeated the two all-time human champions of the TV quiz show *Jeopardy!* on a televised standoff *(more details in Chapter 6).*
Lies (𝒟)	**University at Buffalo:** Researchers trained a system to detect lies with 82 percent accuracy by observing eye movements alone. **Researchers:** Predict deception with 76 percent accuracy within written statements by persons of interest in military base criminal investigations.
Insults (𝒟)	**Imperium:** This data integrity company sponsored a competition to identify insults within online comments and blogs such as "bottom feeder" and "one sick puppy."
Inappropriate comments (𝒟)	**British Broadcasting Co.:** Predicts which comments will be approved for posting on its Web pages so that only one-quarter of the millions of attempted posts need be screened by human moderators.
Sarcasm (𝒟)	**Hebrew University:** Identifies 83 percent of sarcastic Amazon .com product reviews (e.g., "Trees died for this book?").
Dissatisfaction (𝒟)	**PayPal:** Identifies from written feedback customers who intend to leave (aka churn or defect) with 85 percent accuracy. **Citibank:** Categorizes incoming customer messages in order to automatically route problems to the correct support personnel.
Driver inattentiveness (𝒟)	**Ford Motor Co.:** Learned from data to detect when a driver is not alert due to distraction, fatigue, or intoxication. Given an alert driver and a nonalert driver, it can identify the nonalert driver with 86 percent accuracy. **Averitt:** This transportation company predicts truck driver fatigue, crediting this capability with a 30 percent reduction in driver accidents. **Air Force:** Funded research to detect driver fatigue from infrared-illuminated video of the driver.
Psychopathy (𝒟)	**Online Privacy Foundation:** Sponsored a competition to predict psychopathy, as otherwise accessed via nine psychological questions, from the subject's tweets.

What's predicted:	Example organizations that use predictive analytics:
Schizophrenia (𝒟)	**Analytics leaders and a psychiatry professor**: Derived a method to detect patient schizophrenia from the frequent use of pronouns and the brevity of responses to questions, correctly assessing 27+ of 29 previously unseen patients from transcripts alone.
Beauty (𝒟)	**Yahoo! Labs:** Developed a model to categorize photographic portraits as to the subjective human aesthetic of beauty with 64 percent accuracy based on various image attributes. The study determined "that race, gender, and age are largely uncorrelated with photographic beauty."
Brain activity *in order to construct a moving image of whatever you're seeing*	**University of California, Berkeley:** Can render from your brain activities a video approximation of your visual experience. A model predicts the brain activity that will result from what you see so that fMRI readings taken from your brain while viewing a new video can be decoded—it reverse engineers what you're seeing by blending 100 selections from a large video library.
Thoughts (𝒟)	**Researchers:** Computers literally read your mind. Researchers trained systems to decode from fMRI brain scans which type of object you're thinking about—such as tools, buildings, or food—with over 80 percent accuracy for some subjects. **Radica Games:** Manufactures 20Q, a yo-yo-sized toy that employs a neural network to play Twenty Questions, correctly guessing the animal/vegetable/mineral you're thinking of 98 percent of the time after asking you 25 questions; it is robust against wrong answers.

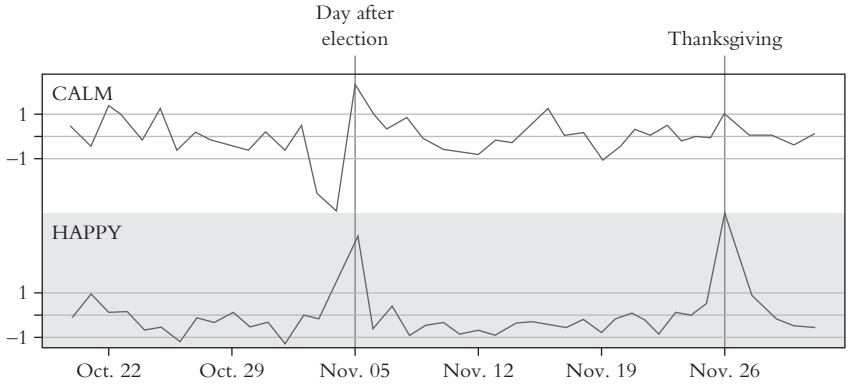
𝒟 Detection, not prediction—see Table 1 for more information.

Table 9 Predictive Analytics in *Workforce: Staff and Employees*

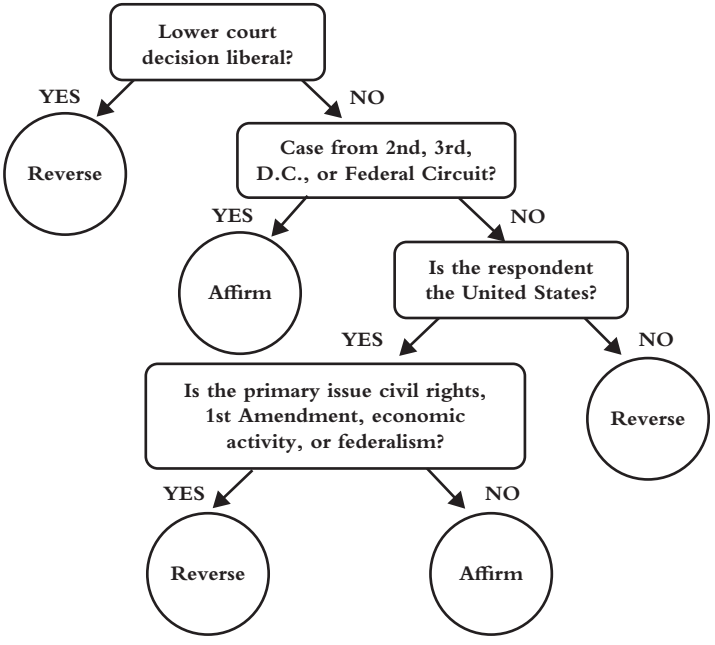
What's predicted:	Example organizations that use predictive analytics:
Quitting	**Hewlett-Packard:** Predictive models generate a "Flight Risk" score for each of its over 300,000 worldwide employees so managers may intervene in advance where possible and plan accordingly otherwise, thus identifying an estimated $300 million in potential savings (*more details in Chapter 2*). **Wikipedia:** Predicts which of its 750,000 editors, who voluntarily perform 139 million edits per year and create over 8,000 new articles a day, will discontinue their work.
Employee longevity	**A midsize U.S. retail bank:** Guided its bank teller hires with predictions as to whether each applicant would last at least 12 months on the job, thus decreasing attrition from 80 percent to 38 percent and saving $600,000 in the first year.
Job performance	**Wells Fargo:** Predicted the most qualified candidates for teller and personal banker positions based on factors such as background experience and skill evaluations, thereby achieving an improvement to employee retention of 12 to 15 percent. **University researchers:** Demonstrated that Facebook profiles predict job performance. Job performance evaluations correlate with personality attributes gleaned from Facebook profiles, such as curiosity, agreeability, and conscientiousness.
Completion of U.S. special forces training	**U.S. Naval Special Warfare Command:** Predicts which candidates will complete the initial phase of training without dropping out (less than 25 percent succeed) in order to support hiring decisions for this highly specialized, demanding job. The number of push-ups a candidate can do is itself a top predictor. However, even if candidates perform poorly in this regard, their success rate is good if they compensate by excelling at running or swimming.
Oil refinery safety incidents	**Shell:** *See entry in Table 6, Fault Detection, Safety, and Logistical Efficiency*
Skills (𝒟)	**LinkedIn:** Labels your profile with skills it predicts you have from your written contents.
Job applications	**CareerBuilder:** Predicts positions for which each job seeker will apply in order to target job recommendations.

𝒟 Detection, not prediction—see Table 1 for more information.

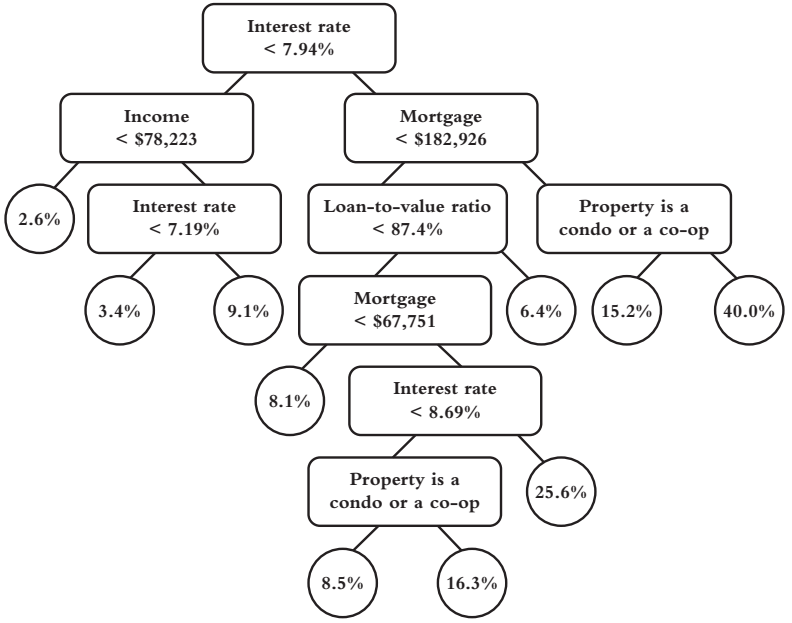
Take a look at how our collective mood moves. Here's sample output of a word-based measure of mood by researchers at Indiana University. Based on a feed from Twitter, it produces daily readings of mass mood for the dimensions *calm* versus *anxious*, and *happy* versus *unhappy* (shown from October 2008 to December 2008):[4]

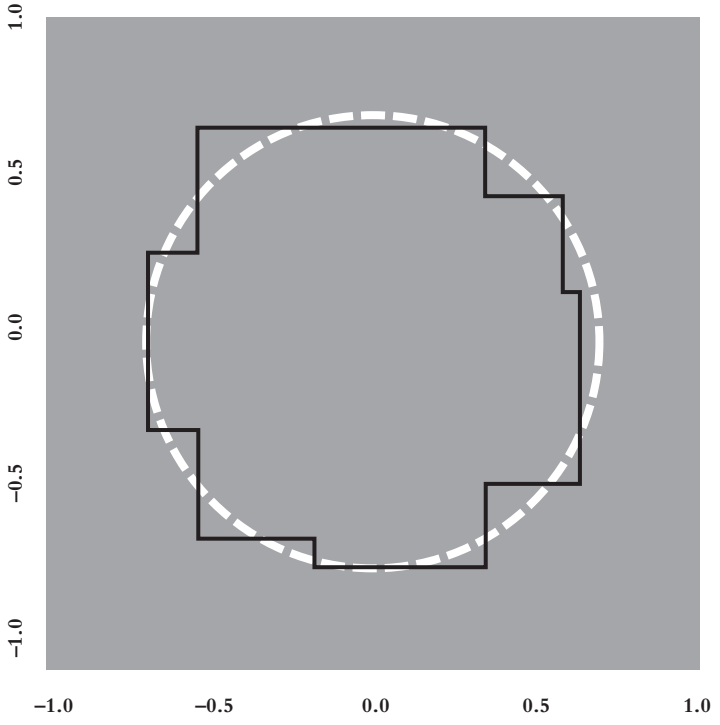


[4] Johan Bollen, Huina Mao, and Xiao-Jun Zeng, "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2, no. 1 (March 2011). Figure reproduced with permission.


Lower court decision liberal?
YES
NO
Reverse
Case from 2nd, 3rd, D.C., or Federal Circuit?
YES
NO
Affirm
Is the respondent the United States?
YES
NO
Is the primary issue civil rights, 1st Amendment, economic activity, or federalism?
Reverse
YES
NO
Reverse
Affirm

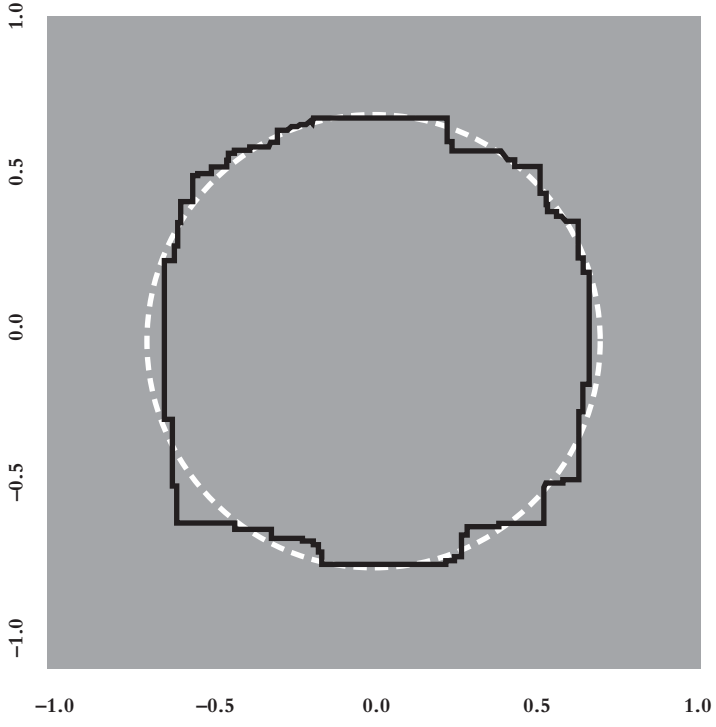
Here's the tree on Chase mortgage data after several more learning steps (this depiction has less annotation—per convention, go left for "yes" and right for "no"):





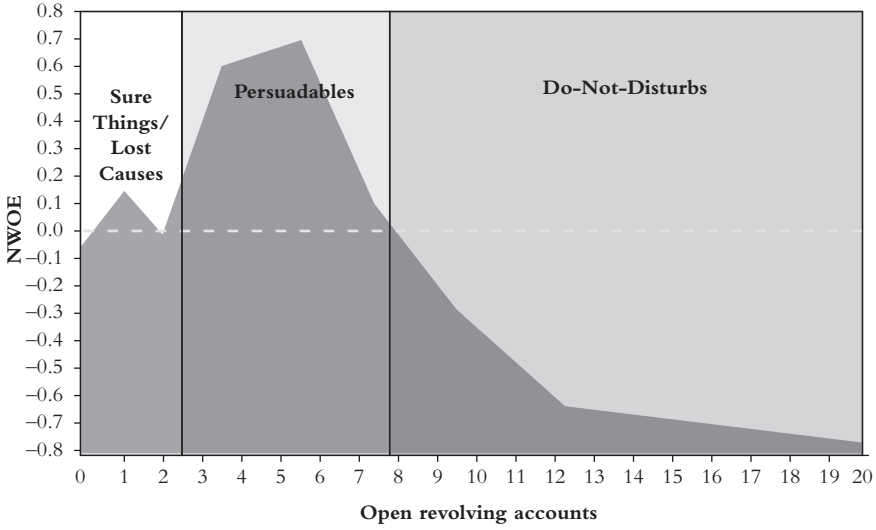
This and the following figure are reproduced with permission.[6]

Bagging a set of 100 CART trees generates a smoother, more refined model:[7]



Reproduced with permission.

[7] While these visuals provide an intuitive view, real PA applications are usually difficult or impossible to view in this way. These examples are two-dimensional, since each case is defined only by the x and y coordinates. Predictive models normally work with dozens or hundreds of variables, in which case the decision boundary cannot be viewed with a two-dimensional diagram. Further, the reality behind the data that the predictive model is attempting to ascertain—in this manufactured example, a single circle—is unknown (if it were known, there would be no need for data analysis in the first place), and is generally more complex than a circle.



Net weight of evidence (NWOE, a measure of uplift) varies by a customer's number of open revolving accounts. Graph courtesy of Kim Larsen.