

Digg Data

- [DATA SCIENCE](#)
- [R-PROJECT](#)
- [PYTHON](#)
- [DIGG BIGG](#)
- [IDEAS](#)

- [Archive](#)
- [RSS](#)
- [Ask me anything](#)

Data Science Use Cases

Background

For each type of analysis think about:

- What problem does it solve and for who?
- How is it being solved today?
- What are the data inputs and where do they come from?
- What are the outputs and how are they consumed- (online algo, static report is a revenue leakage (“saves us money”) or a revenue growth (“makes us money”) problem?)

Use Cases By Function

Sales

- Lead prioritization
 - What is a given lead’s likelihood of closing
 - revenue impact: supports growth
 - usage: online algorithm and static report
- Demand forecasting

Logistics

- Demand forecasting
 - How many of what thing do you need and where will we need them? (Enables lean inventory and prevents out of stock situations.)
 - revenue impact: supports growth and militates against revenue leakage
 - usage: online algorithm and static report

Marketing

- Predicting Lifetime Value (LTV)
 - what for: if you can predict the characteristics of high LTV customers, this supports customer segmentation, identifies upsell opportunities and supports other marketing initiatives

- usage: can be both an online algorithm and a static report showing the characteristics of high LTV customers
- Wallet share estimation
 - working out the proportion of a customer's spend in a category accrues to a company allows that company to identify upsell and cross-sell opportunities
 - Usage: can be both an online algorithm and a static report showing the characteristics of low wallet share customers
 - Churn
 - working out the characteristics of churners allows a company to product adjustments and an online algorithm allows them to reach out to churners
 - usage: can be both an online algorithm and a statistic report showing the characteristics of likely churners
 - Customer segmentation
 - If you can understand qualitatively different customer groups, then we can give them different treatments (perhaps even by different groups in the company). Answers questions like: what makes people buy, stop buying etc
 - usage: static report
 - Product mix
 - What mix of products offers the lowest churn? eg. Giving a combined policy discount for home + auto = low churn
 - usage: online algorithm and static report
 - Cross selling/Recommendation algorithms/
 - Given a customer's past browsing history, purchase history and other characteristics, what are they likely to want to purchase in the future?
 - usage: online algorithm
 - Up selling
 - Given a customer's characteristics, what is the likelihood that they'll upgrade in the future?
 - usage: online algorithm and static report
 - Channel optimization
 - what is the optimal way to reach a customer with certain characteristics?
 - usage: online algorithm and static report

Discount targeting - What is the probability of inducing the desired behavior with a discount - usage: online algorithm and static report

- Reactivation likelihood
 - What is the reactivation likelihood for a given customer
 - usage: online algorithm and static report
- Adwords optimization and ad buying
 - calculating the right price for different keywords/ad slots

Risk

- Credit risk
- Treasury or currency risk
 - How much capital do we need on hand to meet these requirements?
- Fraud detection
 - predicting whether or not a transaction should be blocked because it involves some kind of fraud (eg credit card fraud)
- Accounts Payable Recovery
 - Predicting the probably a liability can be recovered given the characteristics of the borrower and the loan
- Anti-money laundering
 - Using machine learning and fuzzy matching to detect transactions that contradict AML legislation (such as the OFAC list)

Customer support

- Call centers
 - Call routing (ie determining wait times) based on caller id history, time of day, call volumes, products owned, churn risk, LTV, etc.
- Call center message optimization
 - Putting the right data on the operator's screen
- Call center volume forecasting
 - predicting call volume for the purposes of staff rostering

Human Resources

- Resume screening
 - scores resumes based on the outcomes of past job interviews and hires
- Employee churn
 - predicts which employees are most likely to leave
- Training recommendation
 - recommends specific training based of performance review data
- Talent management
 - looking at objective measures of employee success

Use Cases By Vertical

Healthcare

- Claims review prioritization
 - payers picking which claims should be reviewed by manual auditors
- Medicare/medicaid fraud
 - Tackled at the claims processors, EDS is the biggest & uses proprietary tech

- Medical resources allocation
 - Hospital operations management
 - Optimize/predict operating theatre & bed occupancy based on initial patient visits
- Alerting and diagnostics from real-time patient data
 - Embedded devices (productized algos)
 - Exogenous data from devices to create diagnostic reports for doctors
- Prescription compliance
 - Predicting who won't comply with their prescriptions
- Physician attrition
 - Hospitals want to retain Drs who have admitting privileges in multiple hospitals
- Survival analysis
 - Analyse survival statistics for different patient attributes (age, blood type, gender, etc) and treatments
- Medication (dosage) effectiveness
 - Analyse effects of admitting different types and dosage of medication for a disease
- Readmission risk
 - Predict risk of re-admittance based on patient attributes, medical history, diagnose & treatment

Consumer Financial

- Credit card fraud
 - Banks need to prevent, and vendors need to prevent

Retail (FMCG - Fast-moving consumer goods)

- Pricing
 - Optimize per time period, per item, per store
 - Was dominated by Retek, but got purchased by Oracle in 2005. Now Oracle Retail.
 - JDA is also a player (supply chain software)
- Location of new stores
 - Pioneered by Tesco
 - Dominated by [Buxton](#)
- Product layout in stores
 - This is called "plan-o-gramming"
- Merchandizing
 - when to start stocking & discontinuing product lines
- Inventory Management (how many units)
 - In particular, perishable goods
- Shrinkage analytics

- Theft analytics/prevention
(<http://www.internetretailer.com/2004/12/17/retailers-cutting-inventory-shrink-with-spss-predictive-analytic>)
- Warranty Analytics
 - Rates of failure for different components
 - And what are the drivers or parts?
 - What types of customers buying what types of products are likely to actually redeem a warranty?
- Market Basket Analysis
- Cannibalization Analysis
- Next Best Offer Analysis
 - <http://www.analyticbridge.com/xn/detail/2004291:Comment:219197>
- In store traffic patterns (fairly virgin territory)

Insurance

- Claims prediction
 - Might have telemetry data
- Claims handling (accept/deny/audit), managing repairer network (auto body, doctors)
- Price sensitivity
- Investments
- Agent & branch performance
- DM, product mix

Construction

- Contractor performance
 - Identifying contractors who are regularly involved in poor performing products
- Design issue prediction
 - Predicting that a construction project is likely to have issues as early as possible

Life Sciences

- Identifying biomarkers for boxed warnings on marketed products
- Drug/chemical discovery & analysis
- Crunching study results
- Identifying negative responses (monitor social networks for early problems with drugs)
- Diagnostic test development
 - Hardware devices
 - Software

- Diagnostic targeting (CRM)
- Predicting drug demand in different geographies for different products
- Predicting prescription adherence with different approaches to reminding patients
- Putative safety signals
- Social media marketing on competitors, patient perceptions, KOL feedback
- Image analysis or GCMS analysis in a high throughput manner
- Analysis of clinical outcomes to adapt clinical trial design
- COGS optimization
- Leveraging molecule database with metabolic stability data to elucidate new stable structures

Hospitality/Service

- Inventory management/dynamic pricing
- Promos/upgrades/offers
- Table management & reservations
- Workforce management (also applies to lots of verticals)

Electrical grid distribution

- Keep AC frequency as constant as possible
- Seems like a very “online” algorithm

Manufacturing

- Sensor data to look at failures
- Quality management
 - Identifying out-of-bounds manufacturing
 - Visual inspection/computer vision
 - Optimal run speeds
- Demand forecasting/inventory management
- Warranty/pricing

Travel

- Aircraft scheduling
- Seat mgmt, gate mgmt
- Air crew scheduling
- Dynamic pricing
- Customer complain resolution (give points in exchange)
- Call center stuff
- Maintenance optimization

- Tourism forecasting

Agriculture

- Yield management (taking sensor data on soil quality – common in newer John Deere et al truck models and determining what seed varieties, seed spacing to use etc

Mall Operators

- Predicting tenants capacity to pay based on their sales figures, their industry
- Predicting the best tenant for an open vacancy to maximise over all sales at a mall

Education

- Automated essay scoring

Utilities

- Optimise Distribution Network Cost Effectiveness (balance Capital 7 Operating Expenditure)
- Predict Commodity Requirements

Other

- Sentiment analysis
- Loyalty programs
- Sensor data
 - Alerting
 - What's going to fail?
- De duplication
- Procurement

Use Cases That Need Fleshing Out

Procurement

- Negotiation & vendor selection
 - Are we buying from the best producer

Marketing

- Direct Marketing
 - Response rates
 - Segmentations for mailings
 - Reactivation likelihood
 - RFM
 - Discount targeting
 - FinServ
 - Phone marketing
 - Generally as a follow-up to a DM or a churn predictor
 - Email Marketing
- Offline
 - Call to action w/ unique promotion
 - Why are people responding- How do I adjust my buy (where, when, how)?
 - "I'm sure we are wasting half our money here, but the problem is we don't know which ad"
- Media Mix Optimization
 - Kantar Group and Nielson are dominant
 - Hard part of this is getting to the data (good samples & response vars)

Healthcare

- CRM & utilization optimization
- Claims coding
- Formulry determination and pricing
- How do I get you to use my card for auto-pay? Paypal? etc. Unsolved.
- Finance
 - Risk analysis
 - Automating Excel stuff/summary reports

Source: kaggle.com

- [#ideas](#)
- [#datascience](#)
- [#DiggBigg](#)
- [#use-cases](#)
- [2 years ago](#)
- [0 Comments](#)
- [Permalink](#)

Share

Short URL

<http://tumblr.co/Z57mCsISWyj>