# Introduction to Quantitative Consulting

“Reports that say that something hasn't happened are always interesting to me, because as we know, there are **known knowns**; there are things we know that we know. There are **known unknowns**; that is to say, there are things that we now know we don't know. But there are also **unknown unknowns** – there are things we do not know we don't know.”

(Donald Rumsfeld, US Department of Defense News Briefing, 2002)

“Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.”

(attributed to Cliff Stoll in Keeler's *Nothing to Hide: Privacy in the 21st Century*, 2006)

# Contents

# The Consulting Framework

“The perfect consultant is both reliable and extremely skilled.
That said, I will always pick **good and reliable**
over **great but flaky**.”

(Bronwyn Rayfield, paraphrased)

# What is Consulting?

Consulting is the practice of providing an individual or organization with **expertise** in a field in exchange for a **fee**.

Consultants may be hired to **supplement** existing staff (NOT as an employee) or to provide an **external perspective**.

Duties may involve:

- making recommendations to improve products or services
- implementing solutions
- breathing new life into a failing project
- training employees
- re-organizing a company's structure to remove inefficiencies, etc.

# Consulting Framework

## Market Knowledge and Capability

- technical discipline
- sector specialization

## Consulting Competencies

- business understanding and external awareness – PESTLEE (political, economical, social, technological, legal, environment, ethics)
- managing client relationship
- consulting process – engage, develop, deliver, disengage
- consulting tools and methods

## Consulting Skills and Behaviours

- business acumen
- project management
- solid teamwork
- personal and professional development
- professionalism
- ethics
- analytical, predictive, and creative thinking
- emotional intelligence
- effective communication

(International Management Consulting, USA)

# Consultant Types

## Strategy Consultants

- focus on corporate strategy, economic policy, government policy, etc.
- projects for senior managers which are more advisory, less implementation
- quantitative / analytical skills

## Operations Consultants

- focus on improving the performance of operations
- work with both strategy and technology people (in sales, marketing, production, finance, HR, logistics, etc.)
- projects vary from advisory to implementation

# Consultant Types

## Human Resource Consultants

- focus on matters pertaining to human resources or on the workplace culture

## Management (Business) Consultants

- focus on variety of organizational concerns
- catch-all term to describe Strategic, Operational and HR consulting

## Financial and Analytical Advisory Consultants

- focus on financial and analytical matters
- subject matter expertise is paramount (tax law, risk analysis, statistics, etc.)
- quantitative / analytical skills

# Consultant Types

## Information Technology Consultants

- focus on development and application of IT, data analytics, security, etc.
- work on projects, not on business-as-usual activities
- technological / technical skills

## Specialized (Expert) Consulting

- consultants may be brought in for a very specific task
- expertise in a specific field is required

(http://www.consultancy.uk/career/types-of-consultants)

# Ethical Considerations

"We have flown the air like birds
and swum the sea like fishes,
but have yet to learn the simple act
of walking the Earth like brothers."

(Martin Luther King, Jr.)

# The Need for Ethics

When large scale data collection first became possible, there was to some extent a “**Wild West**” mentality to data collection and use. Whatever wasn’t proscribed from a technological perspective was allowed (if not mandatory).

Now, however, professional codes of conduct are being devised for data specialists and other professionals, which outline responsible ways to practice data science – i.e., ways that are **legitimate** rather than fraudulent, as well as **ethical**, rather than unethical.

# The Need for Ethics

Although this puts some **extra** responsibility onto data specialists and quantitative consultants, it also provides them with protection from people who hire them to carry out data science in questionable ways – **they can refuse on the grounds that it is against their professional code of conduct**.

# What Are Ethics?

Broadly speaking, ethics refers to the **study** and **definition** of **right and wrong conduct**.

Ethics may consider what is right or wrong action broadly speaking, or consider how broad ethical principles are appropriately applied in more specific circumstances.

As noted by Richard William Paul and Linda Elder,

> "ethics is not the same as social convention,
> religious beliefs, or laws".

# What Are Ethics?

Influential *Western* ethical theories:

- Kant's **golden rule** (do unto others as you would have them do unto you),
- **consequentialism** (the ends justify the means)
- **utilitarianism** (act in order to maximize positive effect)

Influential *Eastern* ethical theories:

- **Confucianism** (virtue from people and motives, not from outcomes?)
- **Taoism** (case-by-case appropriateness of action determines morality)
- **Buddhism** (harmony and self-restraint to avoid causing harm)

# What Are Ethics?

*Ubuntu* ethical tradition:

- **tension** between individual and universal rights
- **global** context of life
- **solidarity**

Maori *tikanga*:

- connection with **spiritual** realm
- **respect** for all things
- **self-determination**
- **reciprocity**

# Ethics in the Consulting Context

Some examples of data ethics questions:

- **Who**, if anyone, owns data?
- Are there **limits** to how data can be used?
- Are there **value-biases** built into certain analytics?
- Are there categories that should **never** be used in analyzing personal data?
- Should some data be **publicly available** to **all** researchers?

[University of Virginia's Centre for Big Data Ethics, Law, and Policy]

Can these questions always be answered?
Are the answers always the same?

# Ethics in the Consulting Context

A general principle of data analysis is to eschew the **anecdotal** in favor of the **general**; from a purely analytical perspective, focusing on specific observations hides the **full picture**.

However data points are not **solely** marks on paper (or electromagnetic bytes on the cloud).

Decisions made on the basis of data science (security, financial, marketing, etc.) may affect real beings, often from marginalized groups, in **unpredictable ways**.

[OCAP® is a registered trademark of the *First Nations Information Governance Centre* (FNIGC), fnigc.ca/OCAP]

# Ethics in the Consulting Context

First Nations Principles of **OCAP®**:

- **Ownership**
  cultural knowledge, data, and information is owned by First Nations communities
- **Control**
  First Nations communities have the right to control all aspects of research and information management that impact them
- **Access**
  First Nations communities must have access to information and data about themselves no matter where it is held
- **Possession**
  First Nations communities must have physical control of relevant data

“A lapse in ethics can be a conscious choice…
but it can also be negligence.”

(Cathy O’Neil and Rachel Schutt, *Doing Data Science*)

# Guiding Principles

“1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.”

(Isaac Asimov’s 3 Laws of Robotics)

# Best Practices

**“Do No Harm”:** data collected from an individual **should not be used to harm** the individual. This may be an ambiguous goal, but **IT IS CRUCIAL THAT IT BE PART OF YOUR PROCESS**.

**Informed Consent:** covers a wide variety of ethical questions, but mainly:

- Individuals must **agree to the collection and use** of their data
- Individuals must have a **real understanding of what they are consenting to**, and of **possible consequences** for them and others

# Best Practices

**Respect "Privacy":** dearly-held principle. Excessively hard to maintain in the age of constant trawling of the Internet for personal data.

**Keep Data Public:** another aspect of data privacy – some (all? most? any?) data should be kept **public**.

**Opt-In/Opt-Out:** Informed consent requires the ability to **not consent**, i.e., to opt out.

- tacit vs. stated consent

# Best Practices

**Anonymize Data:** removal of identifying fields from the dataset prior to analysis.

**"Let the Data Speak":**

- no cherry picking
- importance of validation (more on this later)
- correlation and causation (more on this later, too)
- repeatability

“And yes, **transparency is also the trick to protecting privacy**, if we empower citizens to notice when neighbors infringe upon it. Isn’t that how you enforce your own privacy in restaurants, where people leave each other alone, because those who stare or listen risk getting caught?”

(David Brin, *The Transparent Society*)

# The Good, the Bad, and the Ugly

We learn to analyze data by dabbling with pedagogical data

In practice, data analysis does not exist in a vacuum

Projects can be classified as **Good**, **Bad** or **Ugly**, either from a **technical** or from an **ethical** standpoint (or both)

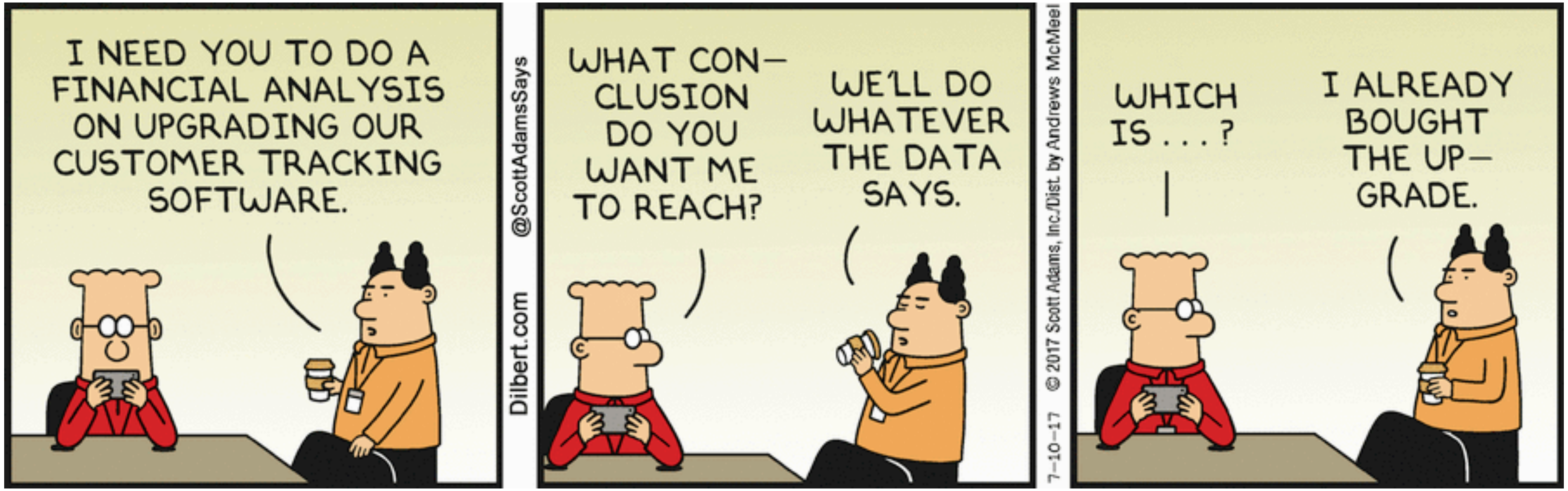
- The **Good:** increases knowledge; can help uncover hidden links, etc.
- The **Bad:** if not done properly, can lead to bad decisions; which can in turn decrease the public's confidence, etc.
- The **Ugly:** there are less savoury applications (to put it mildly)

It is a judgement call – my biases will soon become evident – I will attempt not to get too preachy, but I can’t repeat this enough: **DO NO HARM**!

# Asking the Right Questions

“If we have data, let’s look at data.
If all we have are opinions, let’s go with mine.”

(Jim Barksdale, former Netscape CEO)




[Scott Adams, *Dilbert,* July 12, 2017]

# Asking the Right Questions

Quantitative consulting is about asking & answering questions:

- **Analytics:**

  "How many clicks did this link get?"
- **Data Science:**

  "Based on the previous history of clicks on links of this publisher's site, can I predict how many people from Manitoba will read this specific page in the next three hours?"

  or

  "Is there a relationship between the history of clicks on links and the number of people from Manitoba who will read this specific page?"

# Asking the Right Questions

Quantitative consulting is about asking & answering questions:

- **Quantitative Methods**

  "We have no similar pages whose history could be consulted to make a prediction, but we have reasons to believe that the number of hits will be strongly correlated with the temperature in Winnipeg. Using the weather forecast over the next week, can we predict how many people will access the specific page during that period?"

Data mining models are usually **predictive** (not **explanatory**): they show connections, but don't reveal why the connections exist.

Not every situation calls for analytics,
data science or quantitative methods.

If you can’t ask the right questions,
the project is doomed from the start.

# The Structure of Data

“You can have data without information,
but you cannot have information without data.”

(Daniel Keys Moran, attributed)

# Objects and Attributes

**Object**
- apple

**Shape**
- spherical

**Colour**
- red

**Function**
- food

**Location**
- fridge

**Owner**
- Jen

# From Attributes to Datasets

Attributes are **fields** (or columns) in a database; objects are **instances** (or rows)

Objects are described by their **feature vector**, the collection of attributes associated with value(s) of interest

| ID# | Shape | Colour | Function | Location | Owner |
|---|---|---|---|---|---|
| 1 | spherical | red | food | fridge | Jen |
| 2 | rectangle | brown | food | office | Pat |
| 3 | round | white | tell time | lounge | School |
| ... | ... | ... | ... | ... | ... |

Remember: a person or an object is not simply the sum of its attributes.

# Poisonous Mushroom Dataset

*Amanita muscaria*

**Habitat:** woods

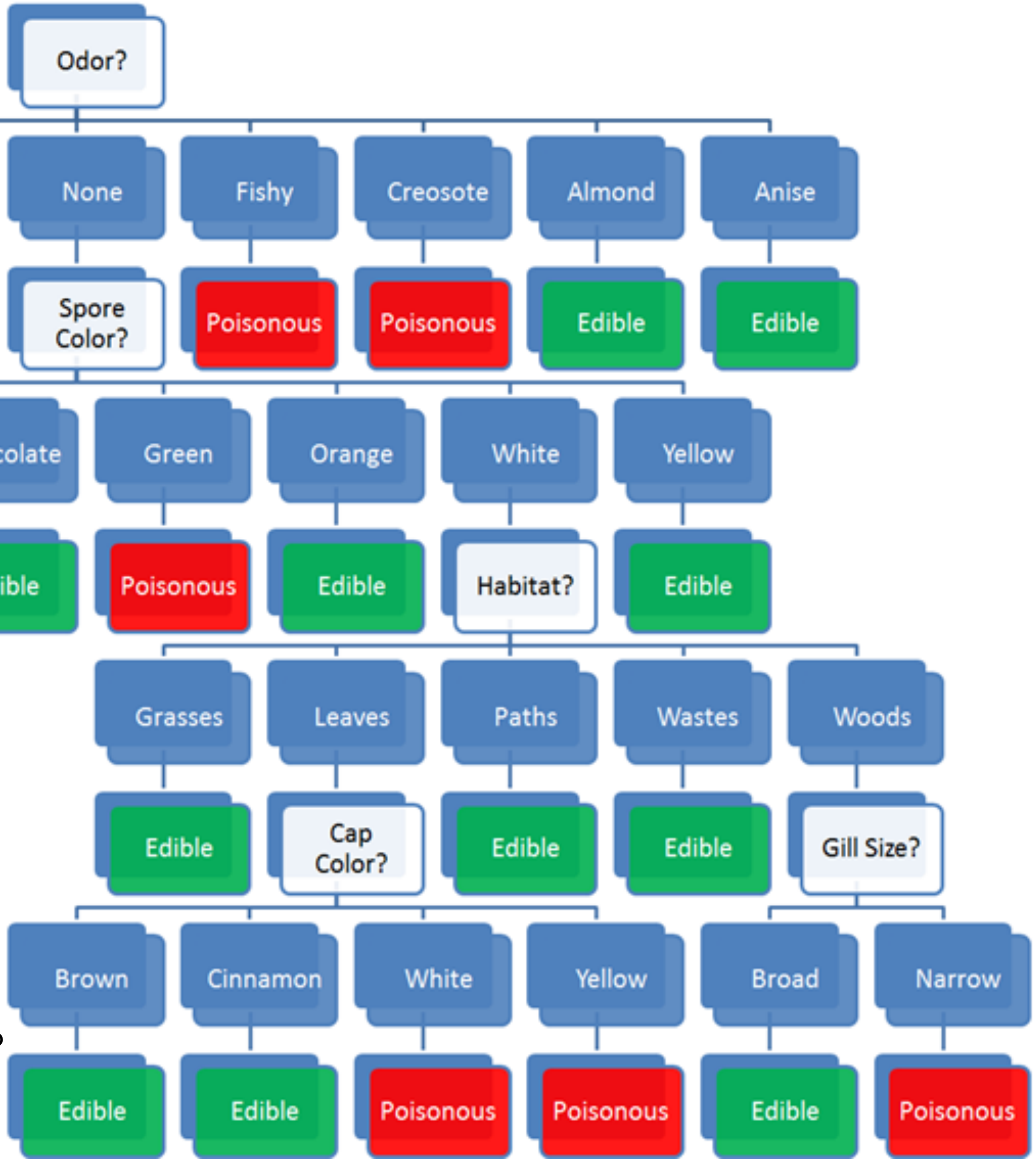
**Gill Size:** narrow

**Odor:** none

**Spores:** white

**Classification problem:**
Is *Amanita muscaria* edible, or poisonous?

[https://archive.ics.uci.edu/ml/datasets/Mushroom]




**Habitat:** woods

**Gill Size:** narrow

**Odor:** none

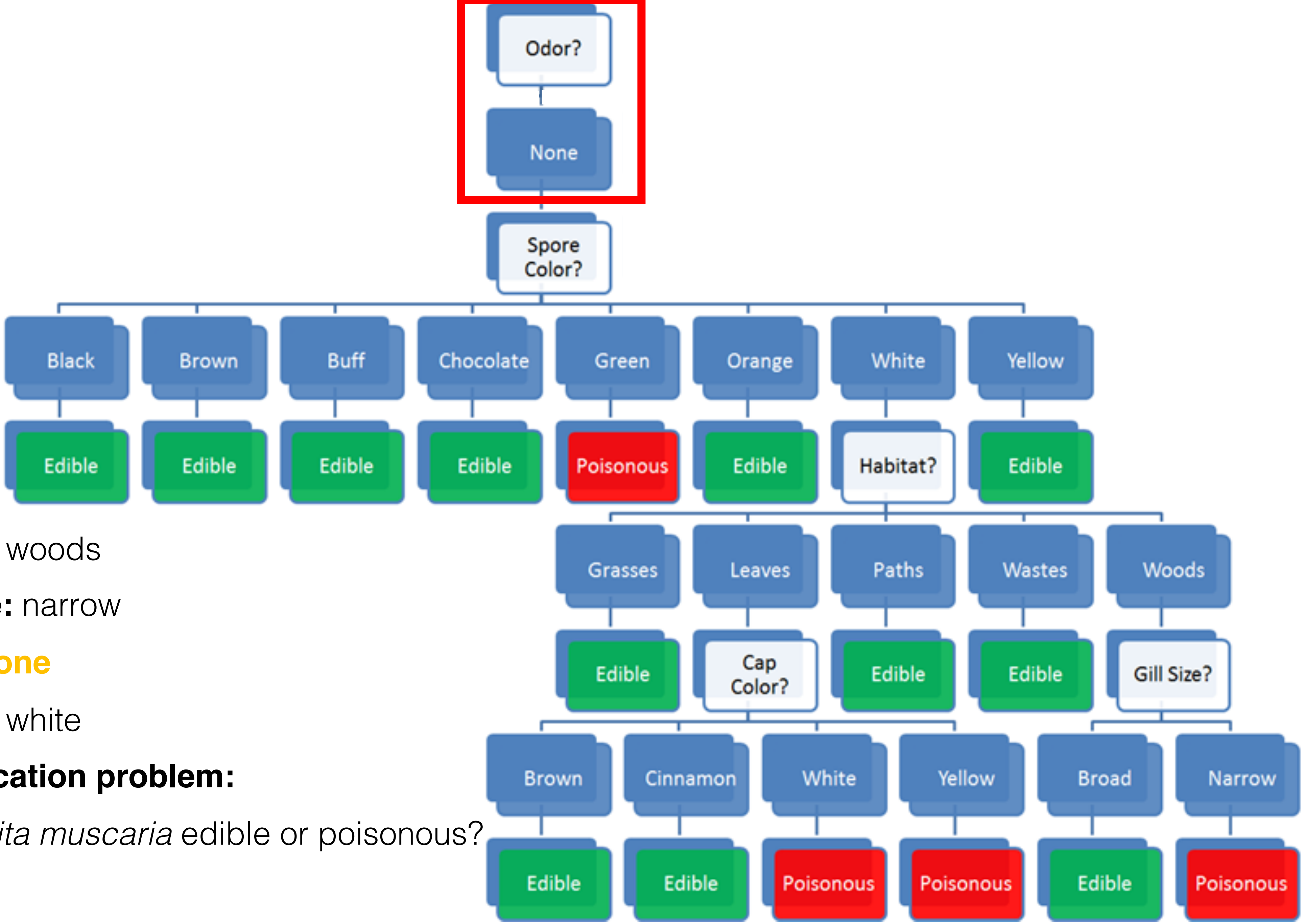
**Spores:** white

**Classification problem:**

Is *Amanita muscaria* edible or poisonous?

[https://archive.ics.uci.edu/ml/datasets/Mushroom]

Where is that model coming from?




**Habitat:** woods

**Gill Size:** narrow

**Odor:** none

**Spores:** white

**Classification problem:**

Is *Amanita muscaria* edible or poisonous?

[https://archive.ics.uci.edu/ml/datasets/Mushroom]

**Habitat:** woods

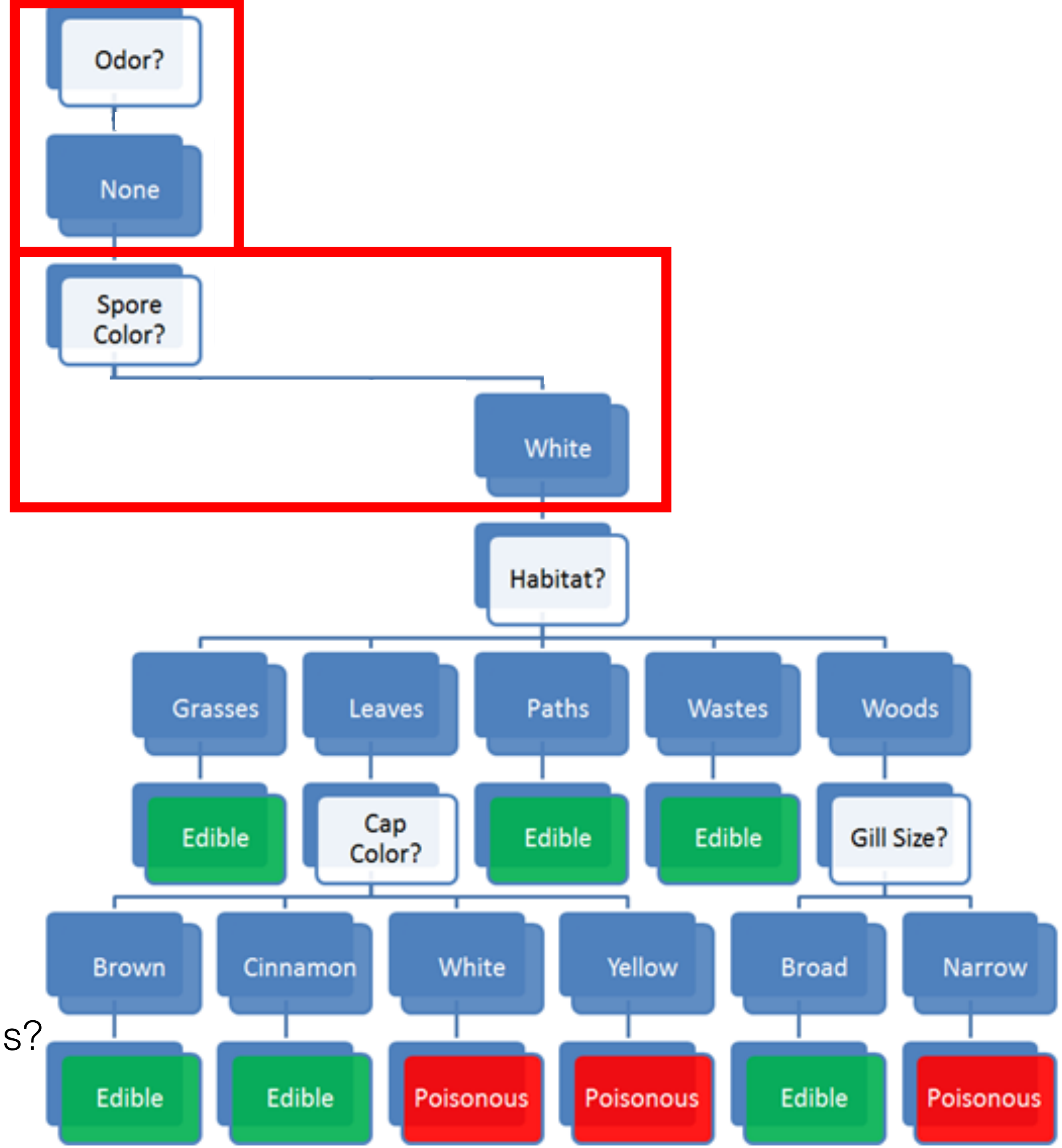
**Gill Size:** narrow

**Odor:** none

**Spores:** white

**Classification problem:**

Is *Amanita muscaria* edible or poisonous?




[https://archive.ics.uci.edu/ml/datasets/Mushroom]

**Habitat:** woods

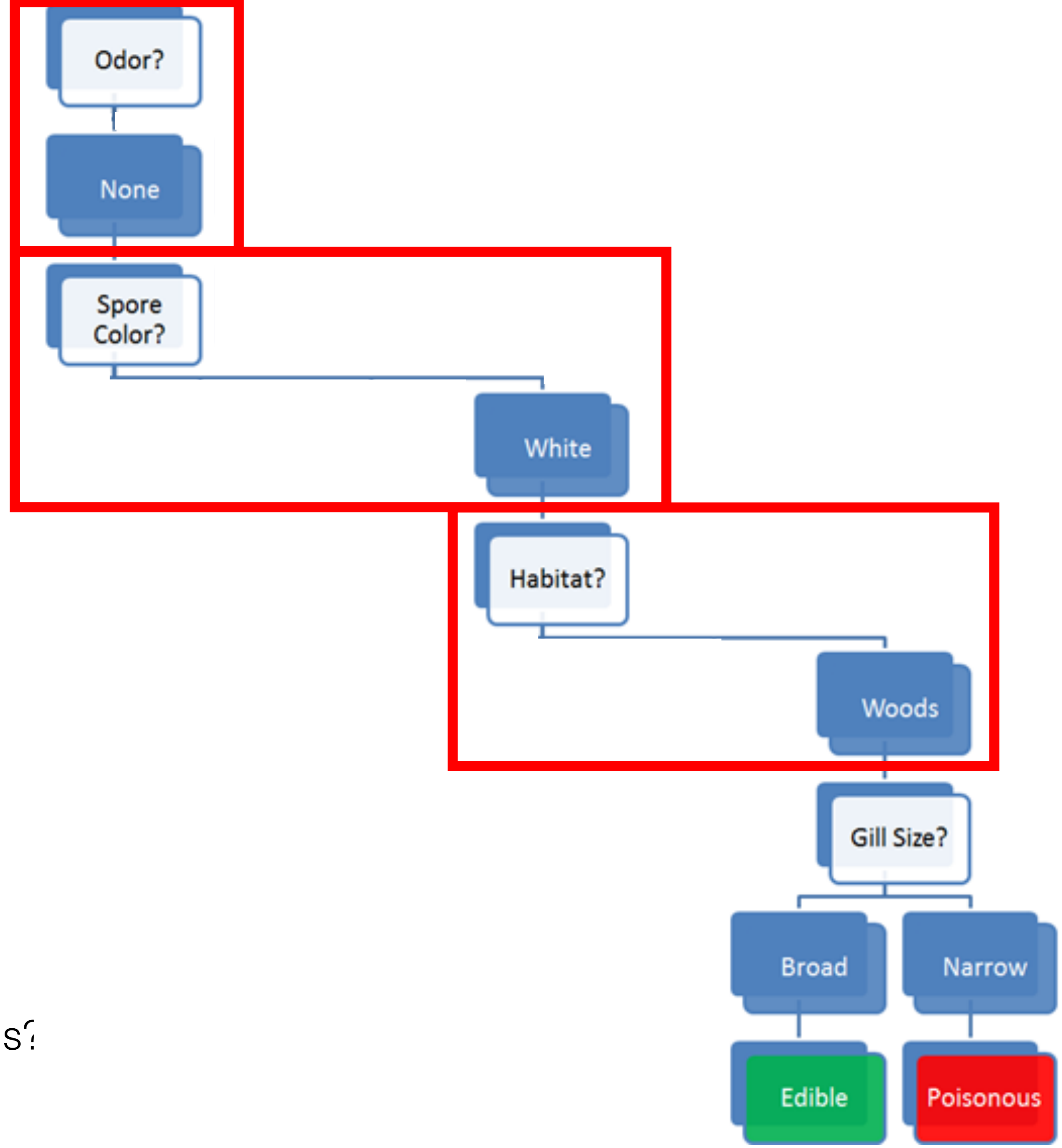
**Gill Size:** narrow

**Odor:** none

**Spores:** white

**Classification problem:**

Is *Amanita muscaria* edible or poisonous?




[https://archive.ics.uci.edu/ml/datasets/Mushroom]

**Habitat:** woods

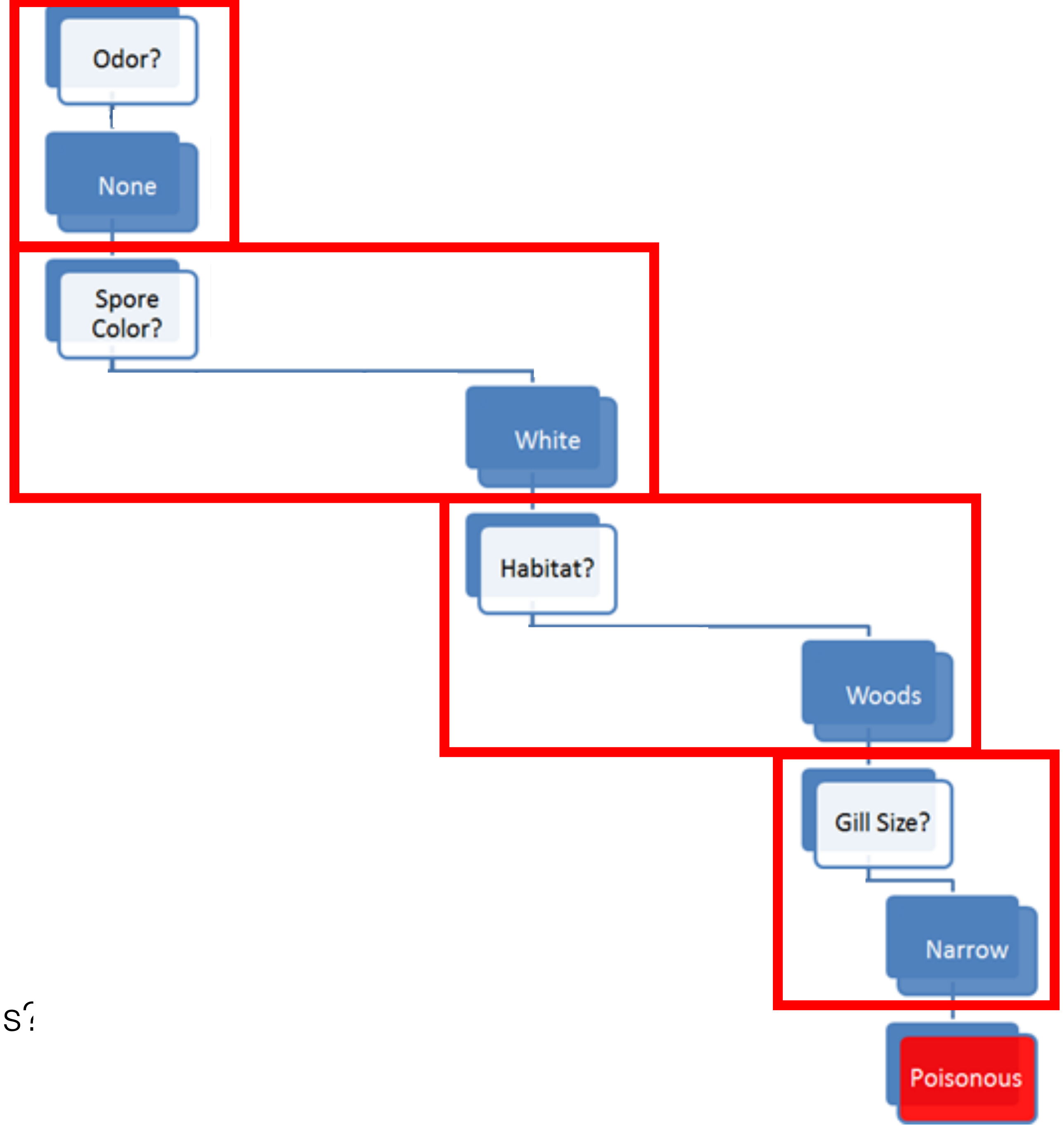
**Gill Size:** **narrow**

**Odor:** none

**Spores:** white

**Classification problem:**

Is *Amanita muscaria* edible or poisonous?




[https://archive.ics.uci.edu/ml/datasets/Mushroom]

**Habitat:** woods

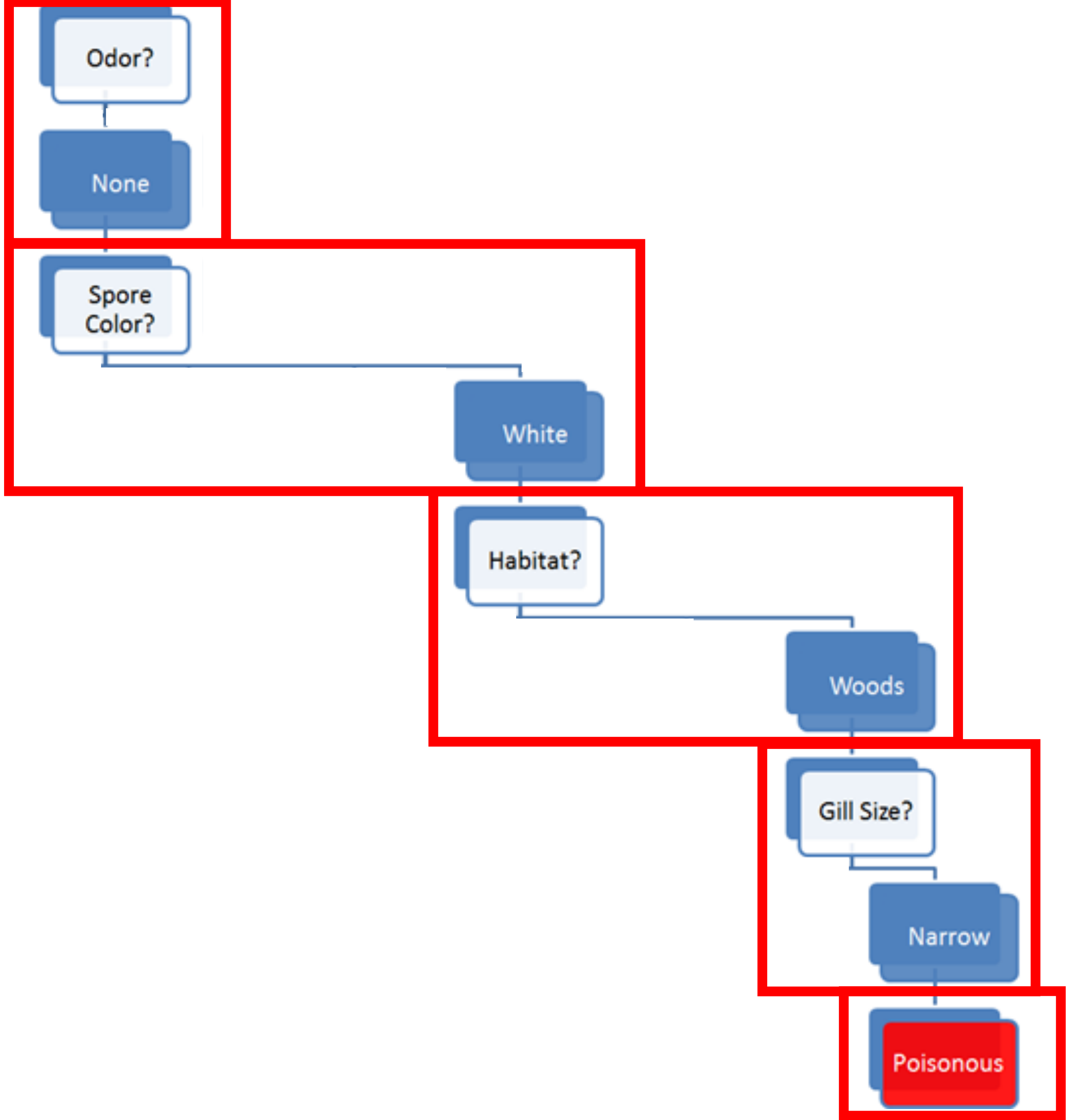
**Gill Size:** narrow

**Odor:** none

**Spores:** white

**Classification problem:**

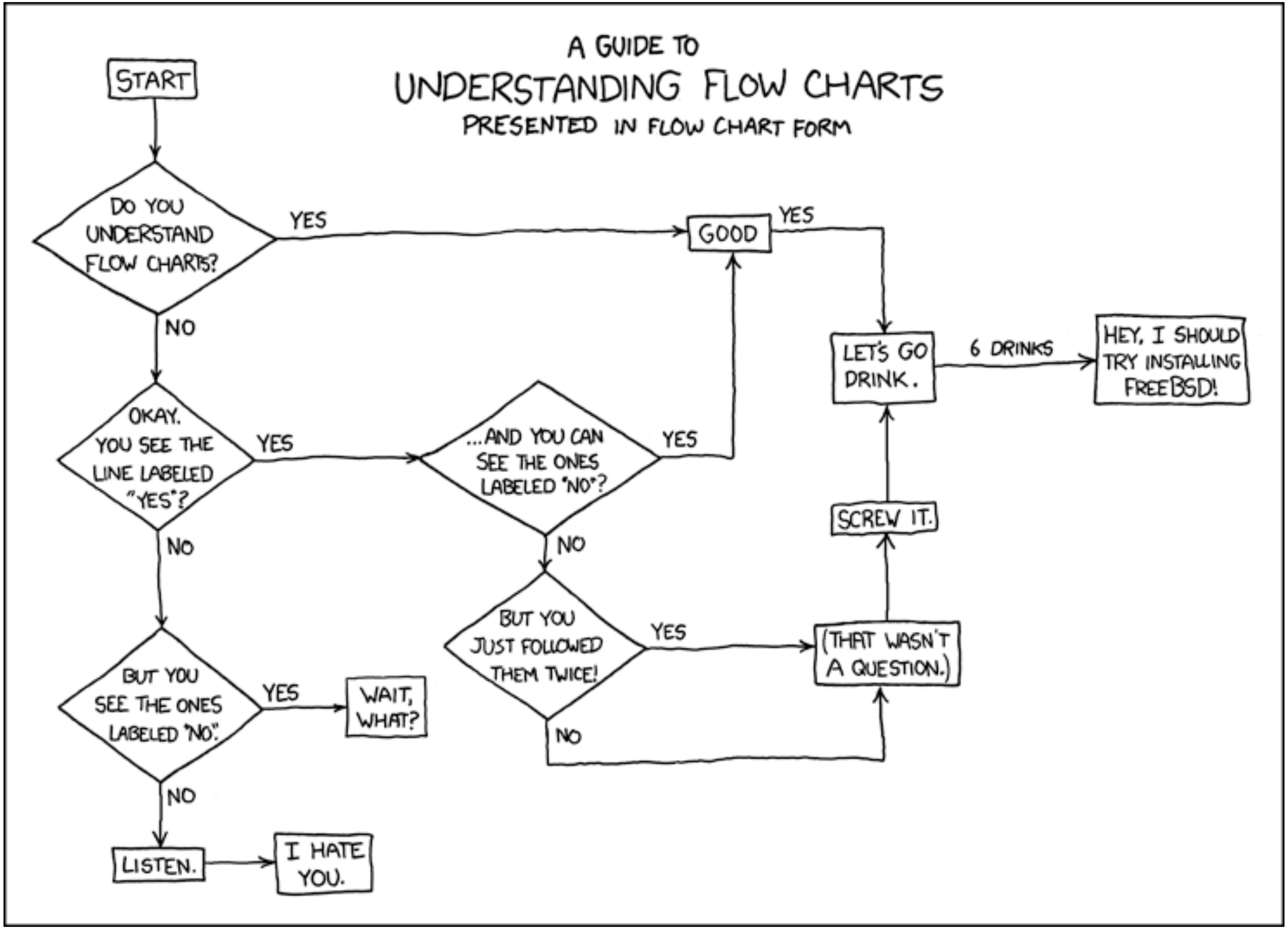
*Amanita muscaria* is **poisonous**




[https://archive.ics.uci.edu/ml/datasets/Mushroom]

What would it take for you to trust an “**edible**” prediction?

# Quantitative Consulting Workflows




[Randall Munroe, xkcd]

# The “Analytical” Method

As with the **scientific method**, there’s a “step-by-step” guide to data analysis:

1. Statement of objective/rationale
2. Infrastructure and data management
3. Data collection
4. Data preparation
5. Data exploration
6. Modeling and analysis
7. Utilization and decision support
8. Communications and documentation

# The “Analytical” Method

Notice that data analysis is a **single** segment of the entire flow.

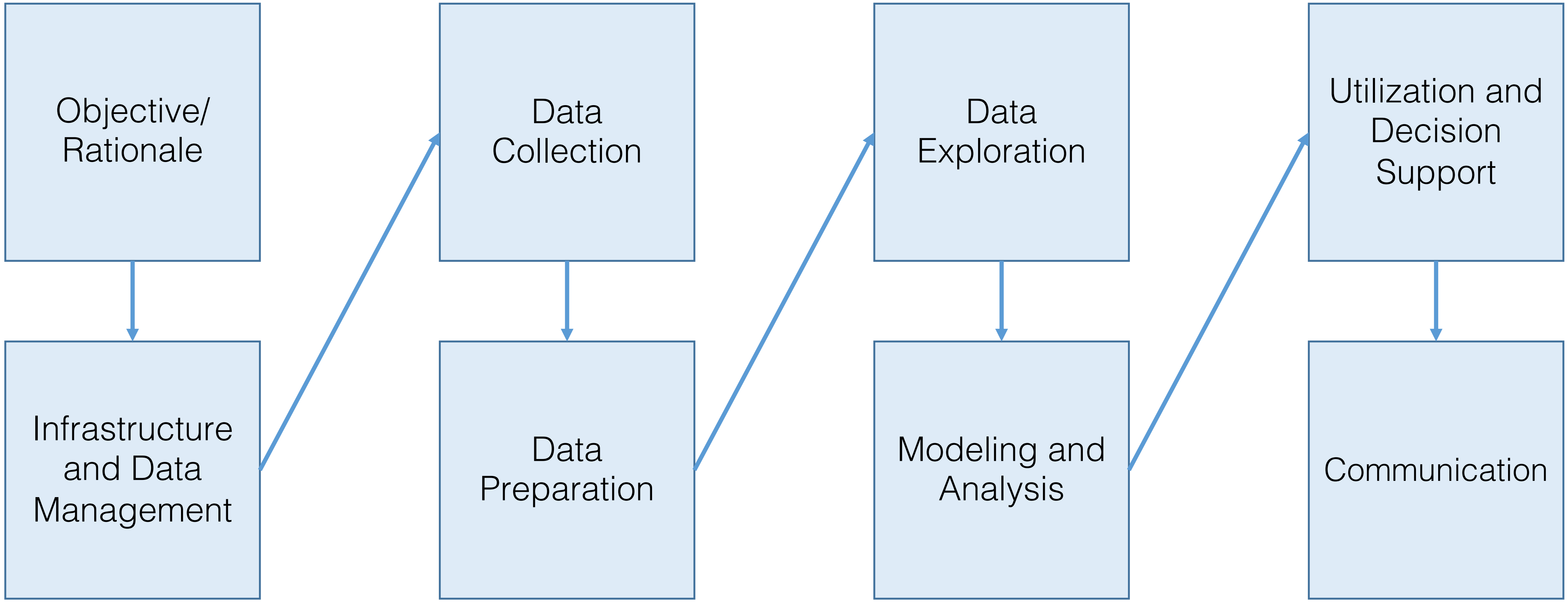
The order is often not respected, and steps are repeated, and it’s a bit of a mess, but it works on the whole (if done correctly).

What about the “**disruptive**” nature of analytic technology?

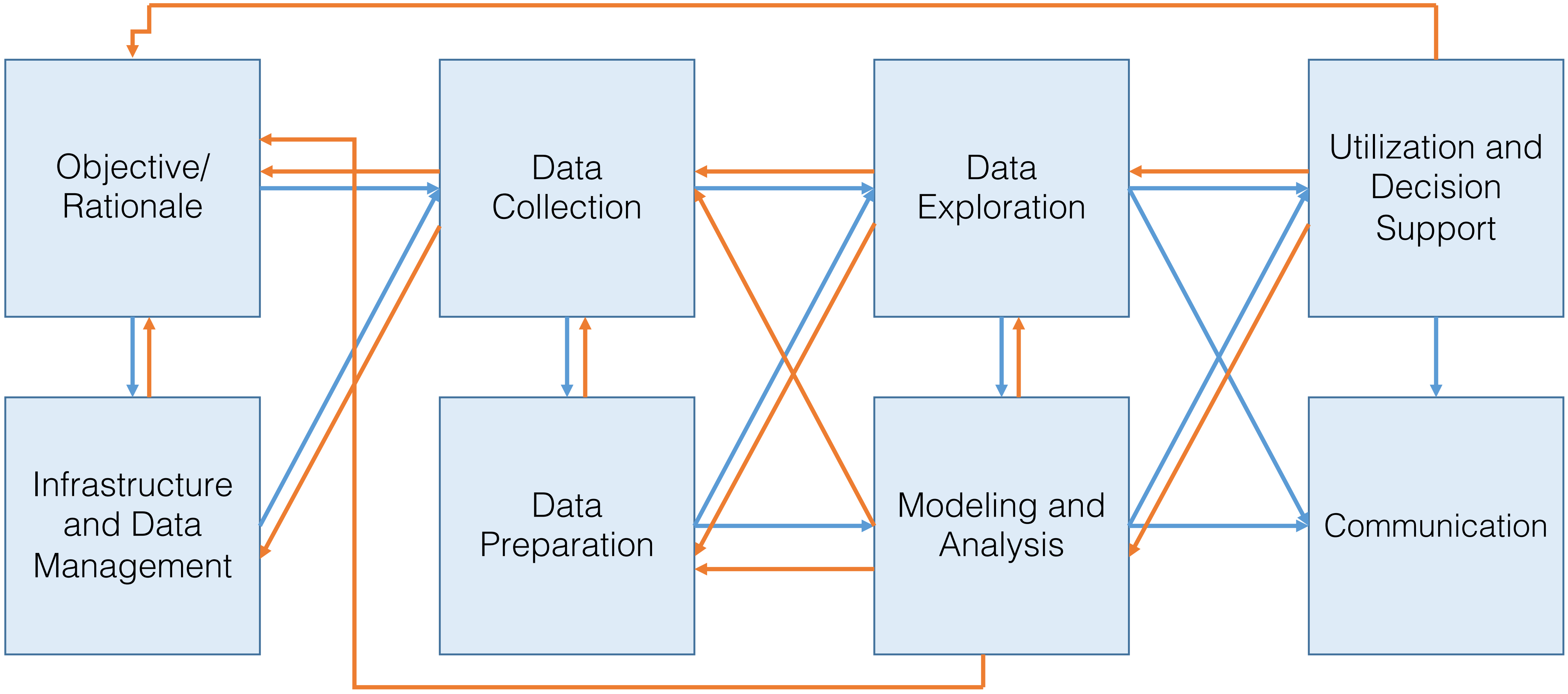
What does it mean to ask a question when the question is to find what’s hidden in the data?

Does it make sense to follow such a structured approach?

# The Data Analysis “Workflow”



# The Data Analysis "Workflow"



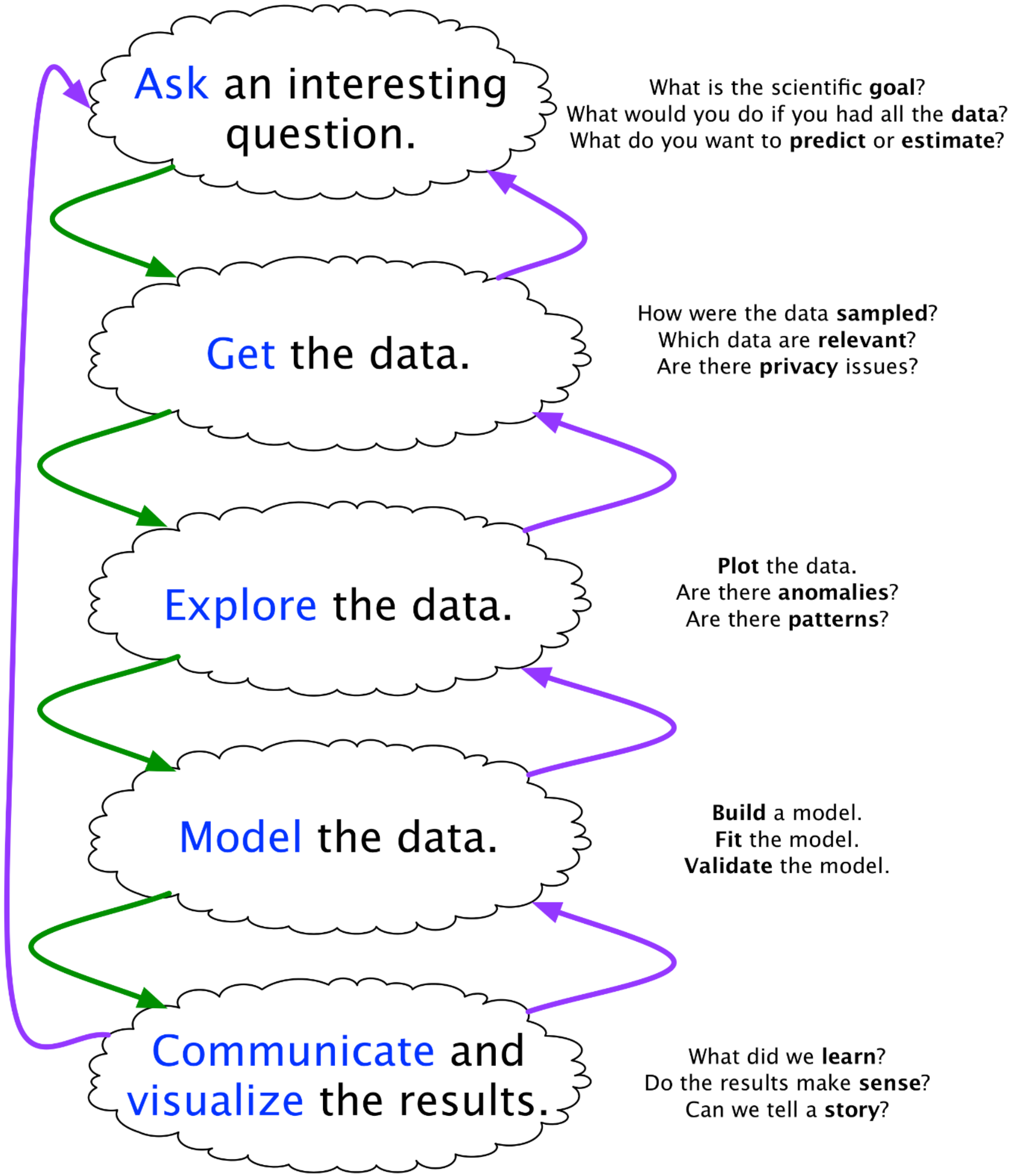
# The Quantitative Consulting Process

A **large number of models** have to be run before a final selection can be made.

**Iterative process:** feature selection and data reduction may require numerous visits to domain experts before models start yielding promising results.

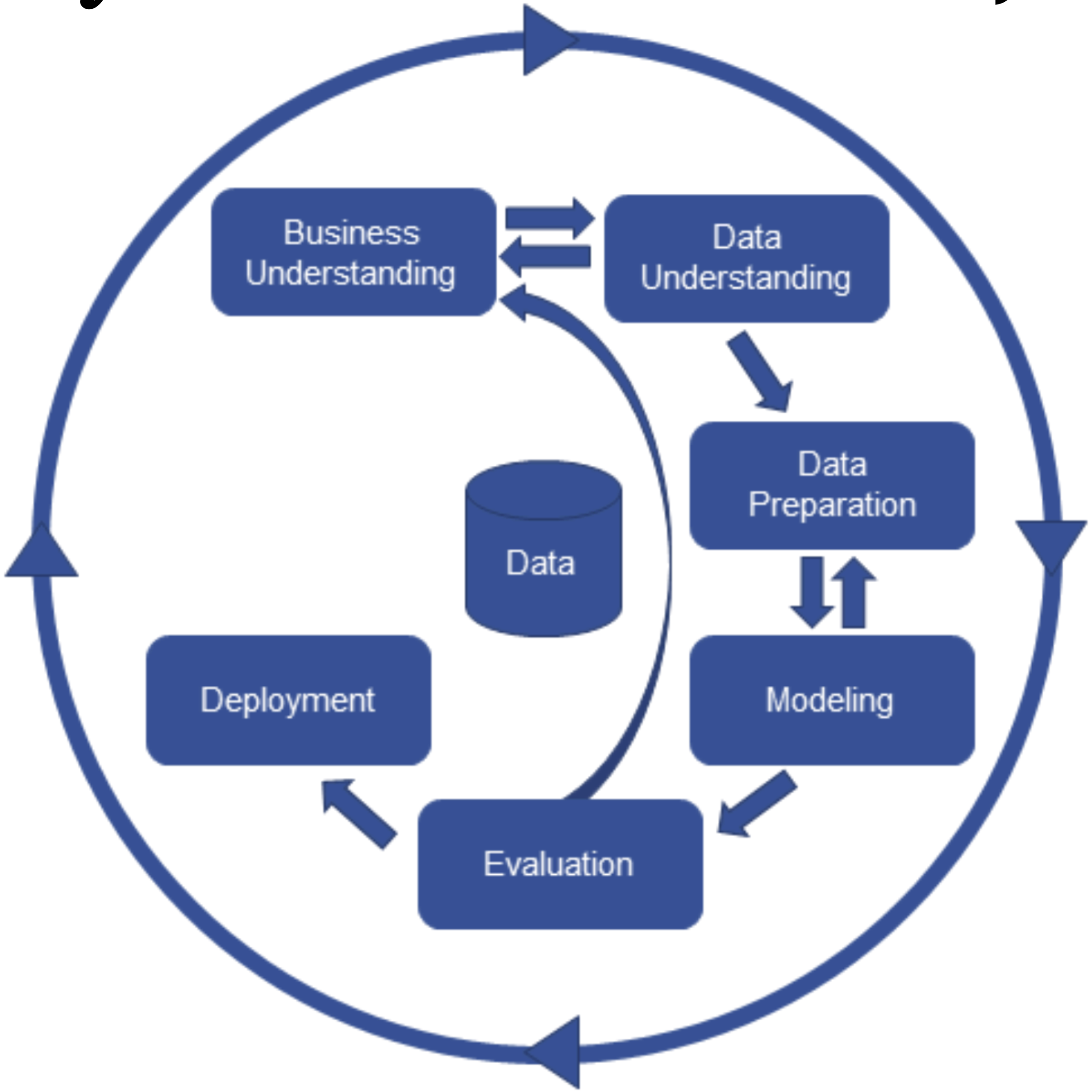
**Domain-specific knowledge** has to be integrated in the models in order to beat random classifiers and clustering schemes, **on average**.

# The Quantitative Consulting Process




[Joe Blitzstein and Hanspeter Pfister, http://cs109.org]

# Cross Industry Standard Process, Data Mining




[James Taylor]

# Cross Industry Standard Process, Data Mining

Understanding the business goal

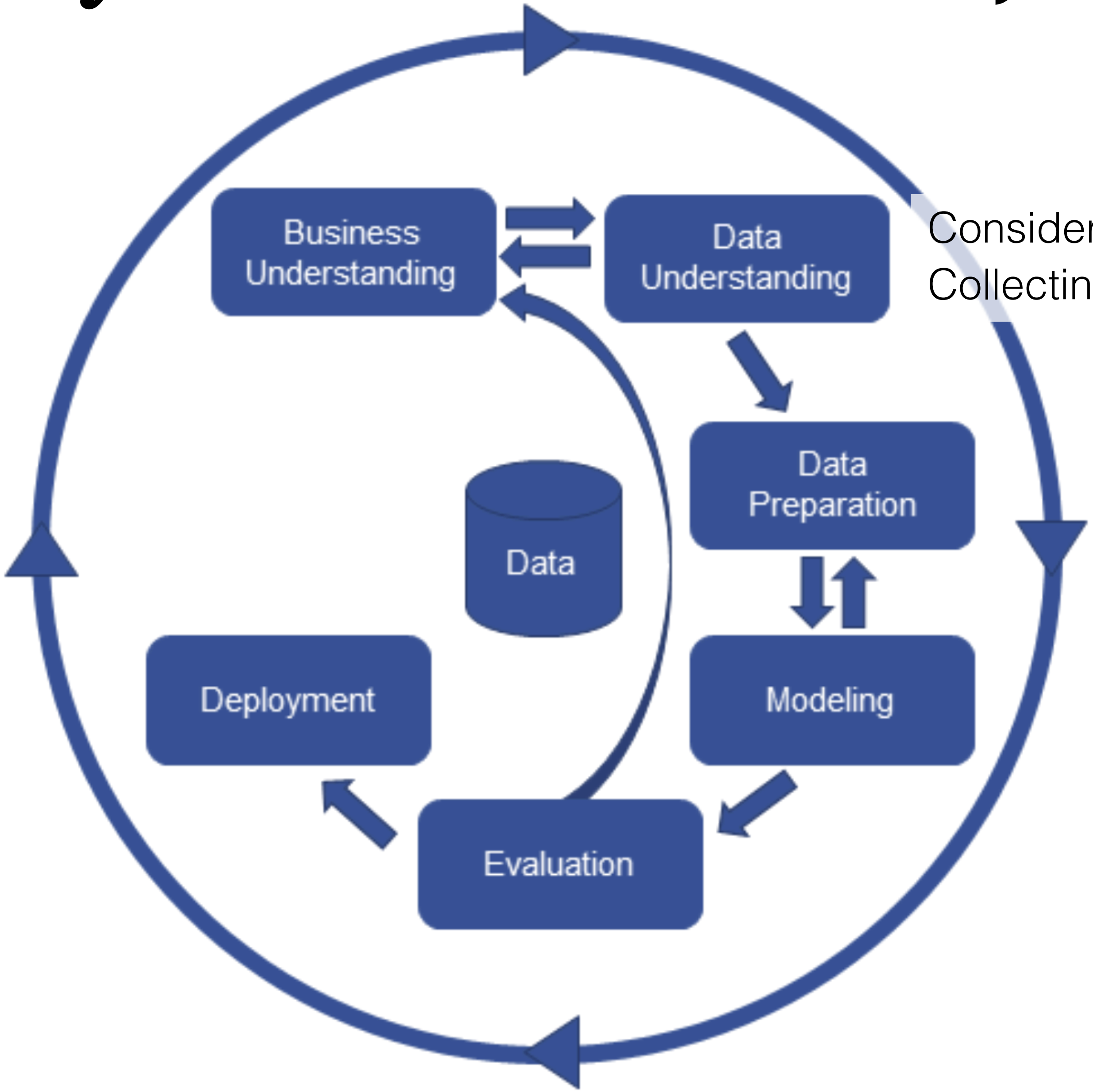
Assessing the situation

Translating the goal in a data analysis objective

Developing a project plan


[James Taylor]

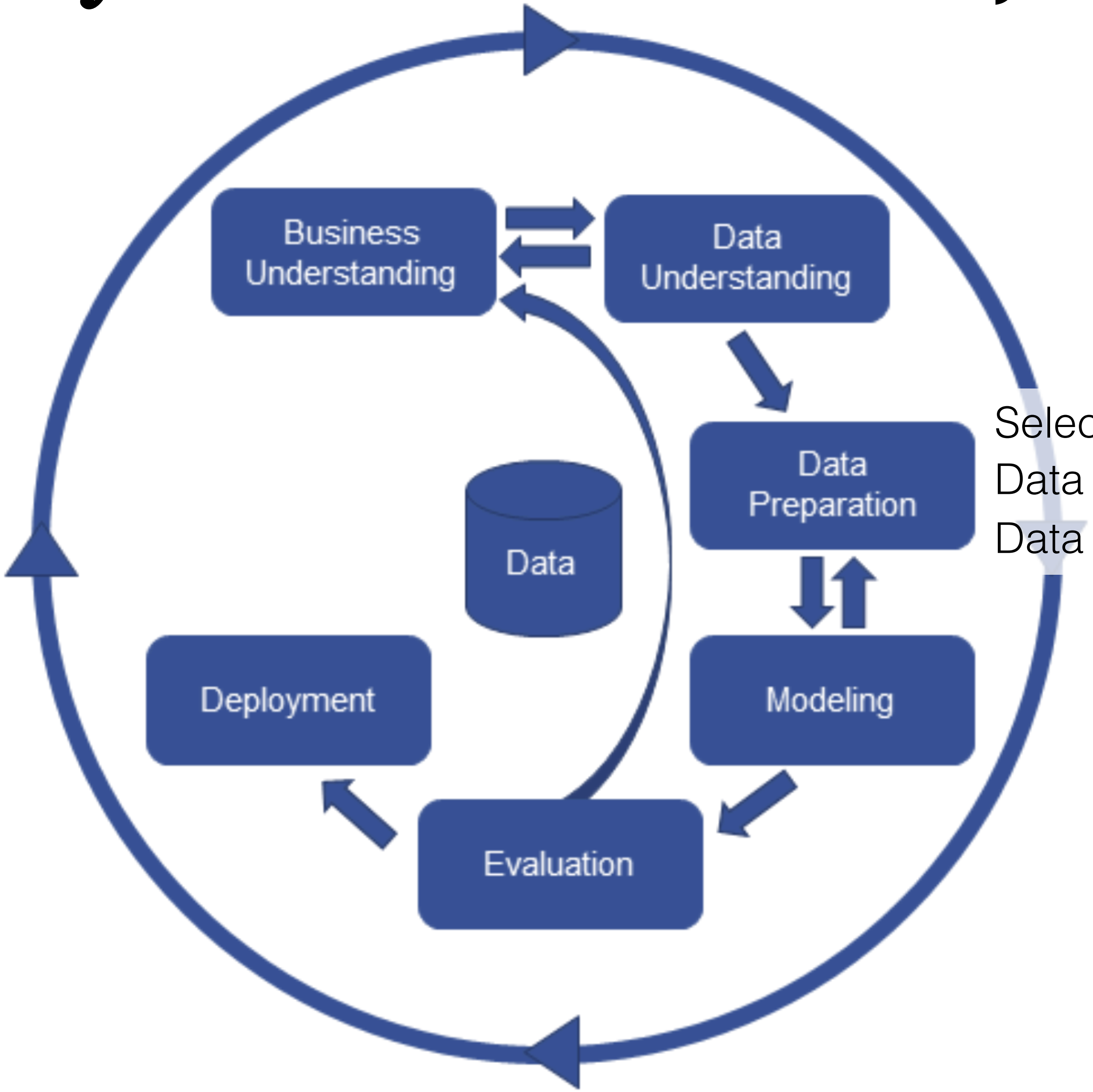
# Cross Industry Standard Process, Data Mining




Considering data requirements
Collecting and exploring data

[James Taylor]

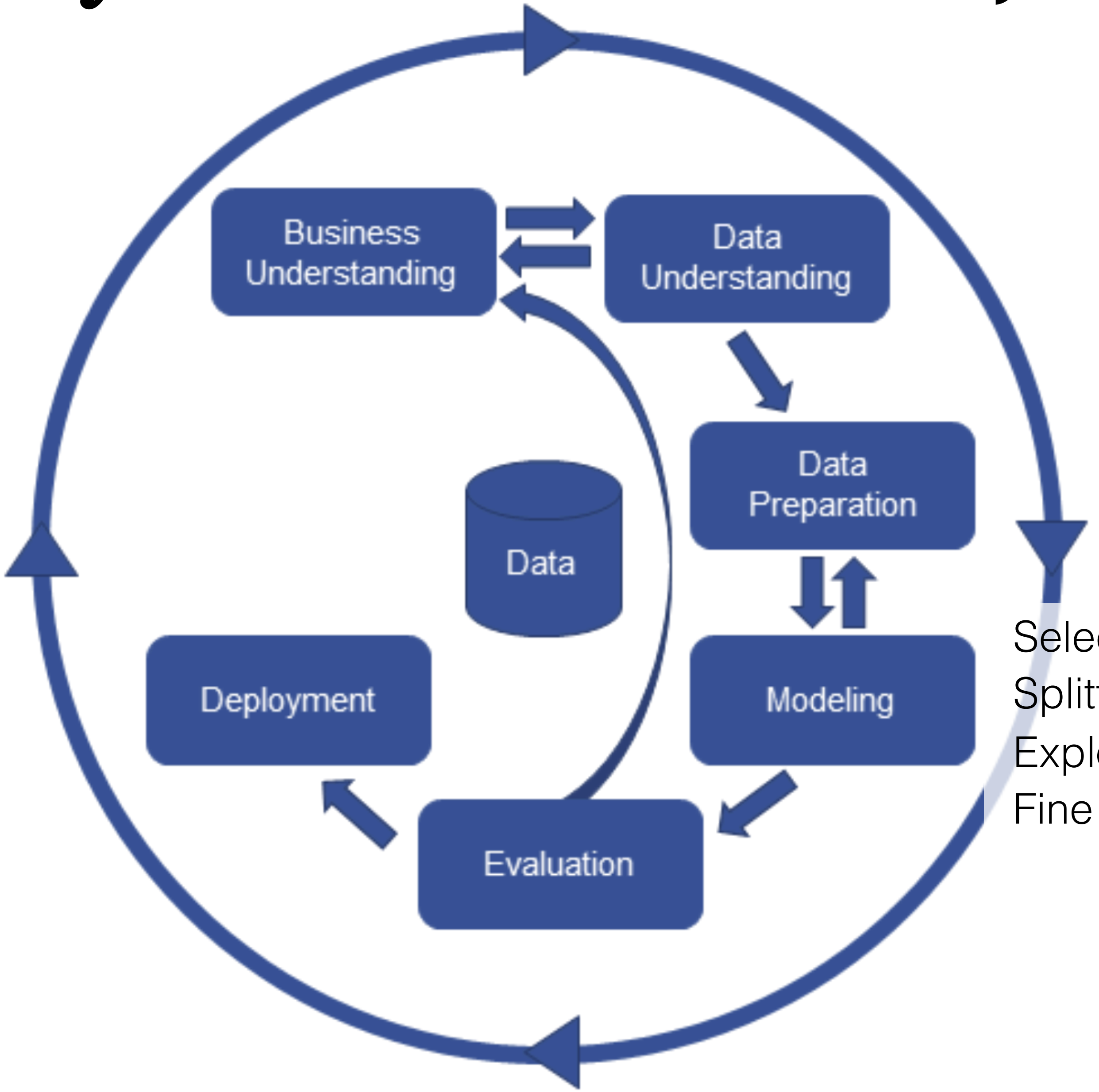
# Cross Industry Standard Process, Data Mining




Selection of appropriate data
Data integration and formatting
Data cleaning and processing

[James Taylor]

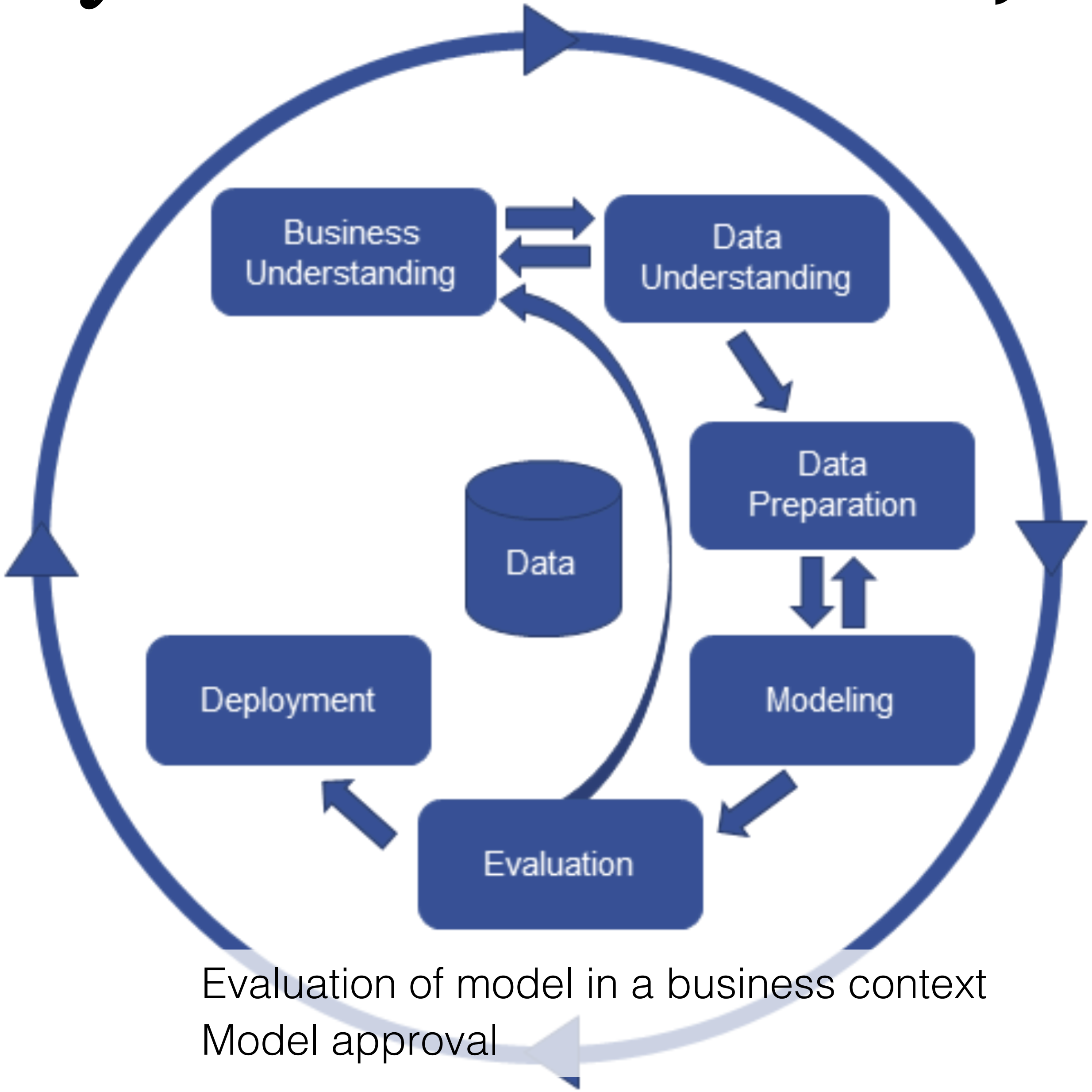
# Cross Industry Standard Process, Data Mining




Selecting appropriate techniques
Splitting into training/testing sets
Exploring alternatives methods
Fine tuning model settings

[James Taylor]

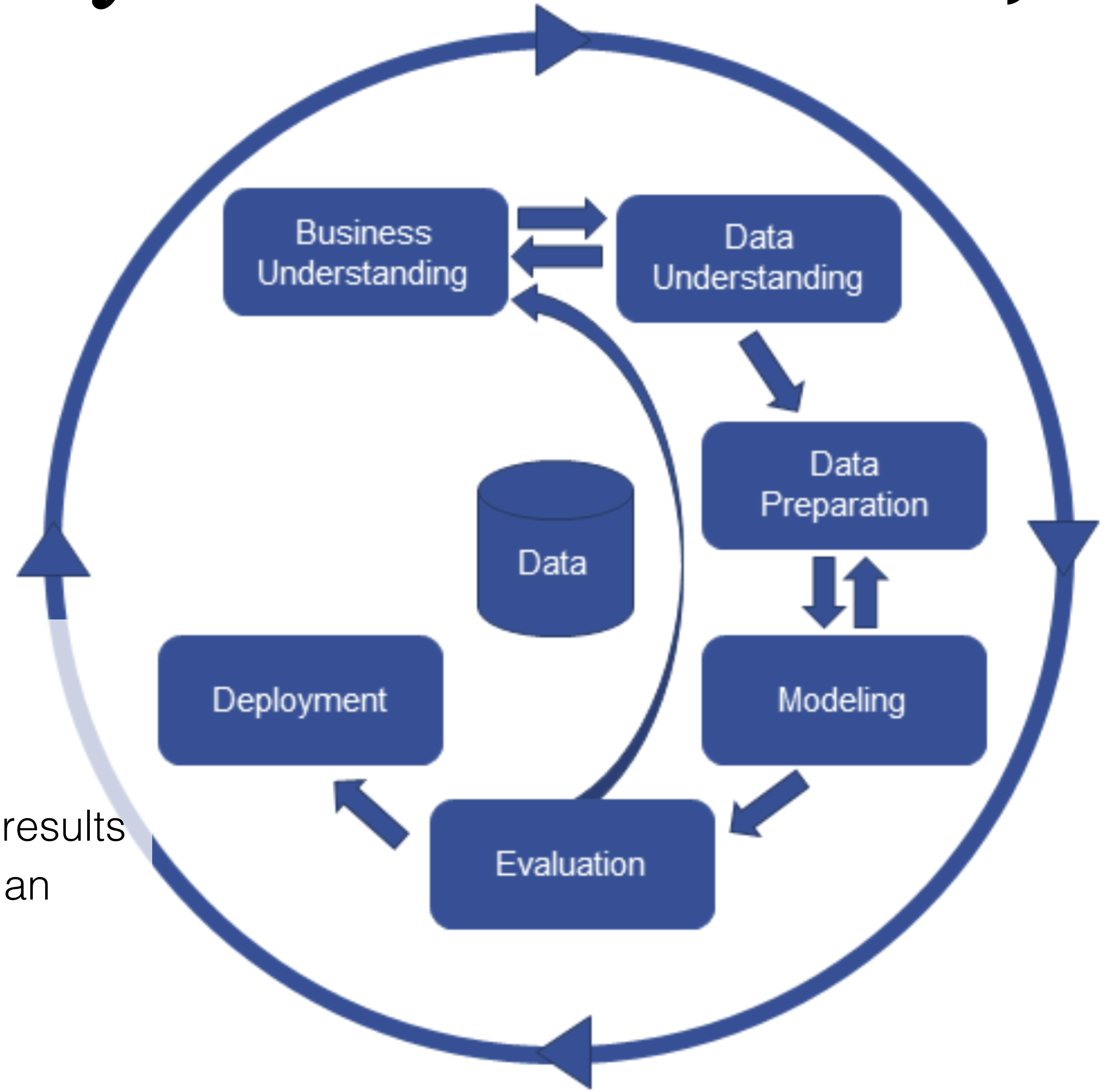
# Cross Industry Standard Process, Data Mining




Evaluation of model in a business context
Model approval

[James Taylor]

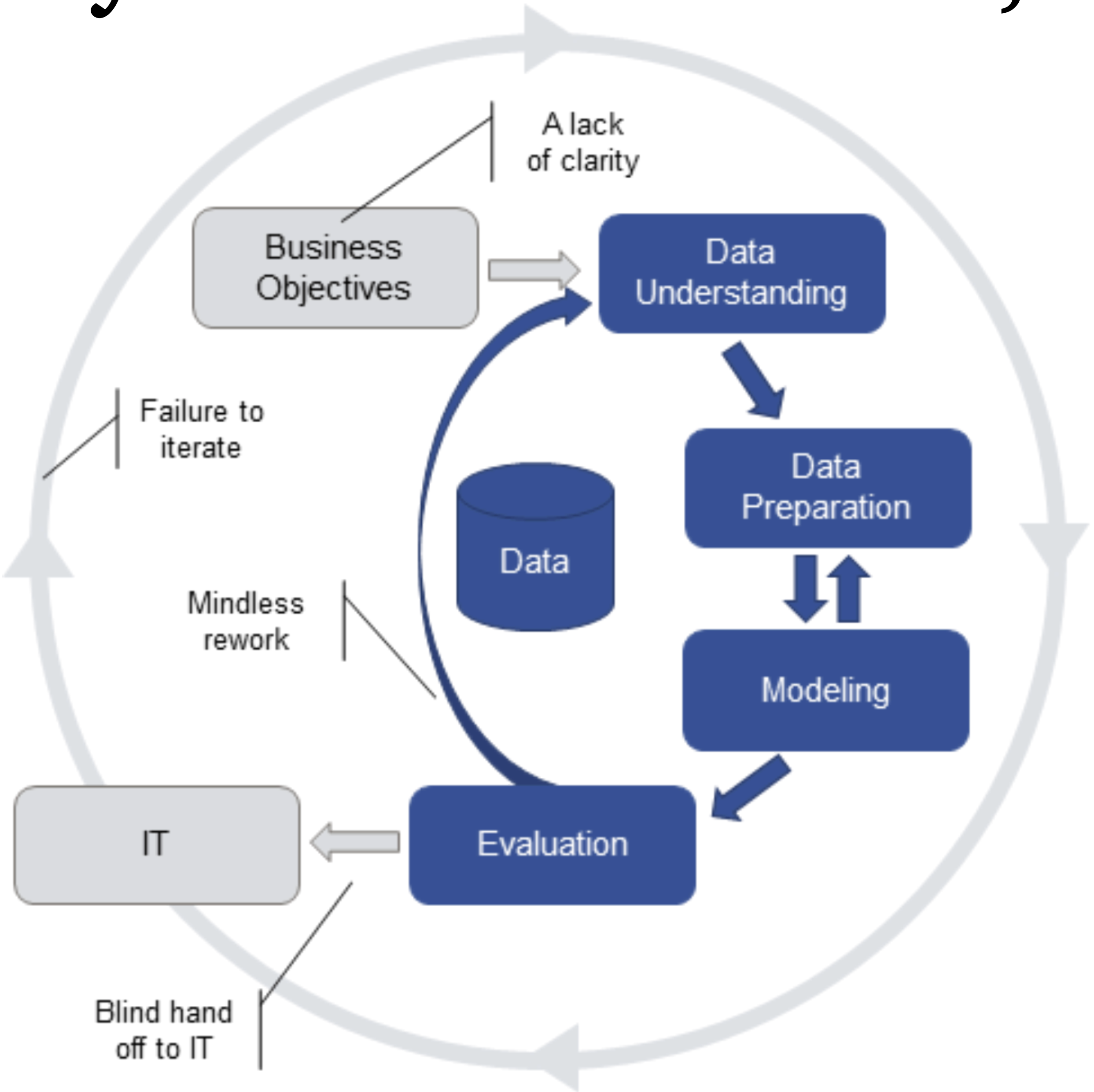
# Cross Industry Standard Process, Data Mining




Reporting findings
Planning the deployment
Deploying the model
Distributing & integrating the results
Developing a maintenance plan
Reviewing the project
Planning the next steps

[James Taylor]

# Cross Industry Standard Process, Data Mining




[James Taylor]

Quantitative consultants AND stakeholders need to keep in mind that the process is both **iterative** and **non-linear**.

# Data Collection and Storage

Data starts to flow in by being **collected**. Some possibilities:

- data collected in only one pass
- data collected in batches
- data collected continuously

Data is **stored**. Data storage (and processing) must reflect:

- how data is being collected (one pass, batch, continuous)
- how much data there is
- type of access and processing that will be required (how fast, how much, by whom)
- stored data may go **stale**

# Data Processing and Modeling

Related to the collection method, data models and metrics (along with other outputs) may be **updated**:

- once
- as new batch data comes in
- continuously as new data comes in

Data can be **processed** and **analyzed** with SEMMA

- **S**ample, **E**xplore, **M**odify, **M**odel, **A**ssess

“All models are wrong. Some models are useful.”

(George Box)

# Model Assessment and Validity

Models should be **current**, **useful**, and **valid**.

Data can be used in conjunction with existing models to come to some conclusions, or can be used to update the model itself.

At what point does one determine that the current data model is **out-of-date**?

- How long does it take Netflix to figure out that you no longer like action movies and want to watch comedies instead?
- How long does it take for Facebook that you are divorced and may not want to be reminded of your ex-partner?

# Model Assessment and Validity

At what point does one determine that the current model is no longer **useful**?

- How long does it take a model to react to a conceptual shift?
- Past successes can lead to **reluctance** to re-assess and re-evaluate a model.

Before applying the findings from a model or an analysis, one must first confirm that the model is reaching **valid conclusions** about the system of interest.

# Life After Analysis

When an analysis or model is ‘released into the wild’, it can take on a life of its own.

Consultants may eventually have to relinquish control over dissemination. Results may be misappropriated, misunderstood, or shelved. What can the analyst do to prevent this?

Finally, because of **analytic decay**, it’s important to view the last analytical step NOT as a static dead end, but rather as an invitation to return to the beginning of the process.

In a very real sense, quantitative consulting
is a never-ending story.

# Roles and Responsibilities

“To leverage Big Data efficiently, an organization needs business analysts, **data scientists**, and big data developers and engineers.”

(De Mauro, Greco, Grimaldi, Ritala)

“A quantitative consultant (singular) is unlikely to get meaningful results – there are simply **too many moving parts** to any data project. Clients need **teams** who understand the challenges faced by their teammates.”

(Boily, Schellinck)

# Quantitative Consulting Ecosystem

Quantitative consulting is a **team sport**, with team members needing a good understanding of both **data** and **context**

- data management
- data preparation
- analysis
- communications

Even slight improvements over a current approach can find a useful place in an organization.


[Meko Deng, 2017]

# Quantitative Consulting Roles

**Managers / Team Leads** have to understand the process to the point of being able to recognize whether what is being done makes sense, and to provide realistic estimates of the time and effort required to complete tasks.

Managers act as **interpreters** between the consulting team and the clients, and **advocate** for the team (and shield them from the client, should the need arises).

They might not be involved with the day-to-day aspects of the projects but are responsible for the project deliverables.

# Quantitative Consulting Roles

**Database Specialists** (DS) work with client and internal IT to ensure that the data sources can be used down the line by the consulting team.

DS **participate** in the analyses, but do not necessarily specialize in esoteric methods and algorithms.

Most quantitative consulting activities require the transfer of some client data to the consulting team. In many instances, this can be as simple as sending a CSV file as an e-mail attachment. In other instances, there are numerous security and size issues.

# Quantitative Consulting Roles

**Technical Specialists** (TS) are team members who work with the processed data to **build sophisticated models** that provide actionable insights.

TS have a **sound understanding** of algorithms and quantitative methods, and of how they can be applied to a variety of data scenarios. They typically have a 2 or 3 areas of expertise and can be counted on to catch up on new material quickly.

TS can also be DS (and vice-versa).

# Quantitative Consulting Roles

**Communication Specialists** are team members who can communicate the actionable insights to managers, policy analysts, decision-makers and other stake holders.

CS **participate** in the analyses, but do not necessarily specialize in esoteric methods and algorithms. CS should keep on top of popular accounts of quantitative results.

CS can also be DS and/or TS (and vice-versa).

# Team Interactions

We can’t get hide it: quantitative consulting is **stressful**.

In an academic environment, the pace is significantly looser:

- deadlines still exist (exams, assignments, theses)
- work can pile up (multiple courses, teaching/TA, etc.)

But the consulting workplace hosts two significant differences:

- a consulting project can only receive 1 of 3 “grades”: **A+** (exceeded expectations), **A-** (met expectation), **F** (didn’t meet expectations)
- while project quality is crucial, so is timeliness – missing a deadline is **just as damaging** as turning in uninspired or flawed work. Perfect work which is delivered late may still cost the client a fair amount of $.

# Team Interactions

Sound **project management** and **scheduling** can help alleviate some of the stress related to these issues.

Quality of **team interactions** can also make or break a project:

- treat colleagues/clients with respect AT ALL TIMES – that includes emails, watercooler conversations, meetings, progress reports, etc.
- keep interactions **cordial** and **friendly** – you don't have to like your teammates, but you're all pulling in the same direction
- keep the team leader/team abreast of **developments** and **hurdles** – delays may affect the project management plan in a crucial manner (plus your colleagues might be able to offer suggestions)
- respond to requests and emails in a **timely manner** (within reason)

# Cognitive Biases

### 1. Anchoring bias.

People are **over-reliant** on the first piece of information they hear. In a salary negotiation, whoever makes the first offer establishes a range of reasonable possibilities in each person's mind.

### 2. Availability heuristic.

People **overestimate the importance** of information that is available to them. A person might argue that smoking is not unhealthy because they know someone who lived to 100 and smoked three packs a day.

### 3. Bandwagon effect.

The probability of one person adopting a belief increases based on the number of people who hold that belief. This is a powerful form of **groupthink** and is reason why meetings are often unproductive.

### 4. Blind-spot bias.

**Failing to recognize** your own cognitive biases is a bias in itself. People notice cognitive and motivational biases much more in others than in themselves.

### 5. Choice-supportive bias.

When you choose something, you tend to feel positive about it, even if that **choice has flaws**. Like how you think your dog is awesome — even if it bites people every once in a while.

### 6. Clustering illusion.

This is the tendency to **see patterns in random events**. It is key to various gambling fallacies, like the idea that red is more or less likely to turn up on a roulette table after a string of reds.

### 7. Confirmation bias.

We tend to listen only to information that confirms our **preconceptions** — one of the many reasons it's so hard to have an intelligent conversation about climate change.

### 8. Conservatism bias.

Where people favor prior evidence over new evidence or information that has emerged. People were **slow to accept** that the Earth was round because they maintained their earlier understanding that the planet was flat.

### 9. Information bias.

The tendency to **seek information when it does not affect action**. More information is not always better. With less information, people can often make more accurate predictions.

### 10. Ostrich effect.

The decision to **ignore dangerous or negative information** by "burying" one's head in the sand, like an ostrich. Research suggests that investors check the value of their holdings significantly less often during bad markets.

[20 Cognitive Biases That Screw Up Your Decisions, *Business Insider*]

# Cognitive Biases

### 11. Outcome bias.

Judging a decision based on the **outcome** — rather than how exactly the decision was made in the moment. Just because you won a lot in Vegas doesn't mean gambling your money was a smart decision.

### 12. Overconfidence.

Some of us are **too confident about our abilities**, and this causes us to take greater risks in our daily lives. Experts are more prone to this bias than laypeople, since they are more convinced that they are right.

### 13. Placebo effect.

When **simply believing** that something will have a certain effect on you causes it to have that effect. In medicine, people given fake pills often experience the same physiological effects as people given the real thing.

### 14. Pro-innovation bias.

When a proponent of an innovation tends to **overvalue its usefulness** and undervalue its limitations. Sound familiar, Silicon Valley?

### 15. Recency.

The tendency to weigh the **latest information** more heavily than older data. Investors often think the market will always look the way it looks today and make unwise decisions.

### 16. Salience.

Our tendency to focus on the **most easily recognizable features** of a person or concept. When you think about dying, you might worry about being mauled by a lion, as opposed to what is statistically more likely, like dying in a car accident.

### 17. Selective perception.

Allowing our expectations to **influence how we perceive** the world. An experiment involving a football game between students from two universities showed that one team saw the opposing team commit more infractions.

### 18. Stereotyping.

Expecting a group or person to have certain qualities without having real information about the person. It allows us to quickly identify strangers as friends or enemies, but people tend to **overuse and abuse** it.

### 19. Survivorship bias.

An error that comes from focusing only on surviving examples, causing us to **misjudge a situation**. For instance, we might think that being an entrepreneur is easy because we haven't heard of all those who failed.

### 20. Zero-risk bias.

Sociologists have found that **we love certainty** — even if it's counterproductive. Eliminating risk entirely means there is no chance of harm being caused.

[20 Cognitive Biases That Screw Up Your Decisions, *Business Insider*]

# Consulting Cheat Sheet

# TL;DR

1. Business solutions are not the same as academic solutions.

2. The data and models don’t always support the client’s hopes, wants, and needs.

3. Communication is key – with the client and with your team.

4. Consultants need to be flexible (within reason) and willing & able to learn something new, quickly.

# TL;DR

5. Not every problem calls for quantitative methods.

6. Learn from your experiences (good and bad).

7. Manage your projects and expectations.

8. Maintain a healthy work-life balance.

# TL;DR

9. Respect the client, the project, the methods, and your team.

10. Remember the analogy between consulting and dating.

11. Don't work without a contract (or a lawyer).

12. Don't oversell, but don't sell yourself short.

# TL;DR

13. Consulting is not about how smart you are, it's about how you can help the client.

14. If you're going to work *pro bono*, do it because you believe in the project, not because you need exposure or experience.

15. When what the client wants can’t be done, offer alternatives.

16. There ain’t no such thing as a free lunch.

# References

# References


1. First Nations – OCAP
2. Wikipedia article on Semi-Supervised Learning
3. Wikipedia article on Supervised Learning
4. Wikipedia article on Reinforcement Learning
5. Wikipedia article on Unsupervised Learning
6. J. Blitzen [2017], *What is it like to design a data science class?*, answer on *Quora*
7. J. Taylor [2017], *4 Problems with CRISP-DM*, KDNuggets.
8. Brin, D. [1998], *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?*, Perseus.
9. Mayer-Schönberger, V. and Cukier, K. [2013], *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Houghton Mifflin Harcourt.

# References

10. Mayer-Schönberger, V. [2009], *Delete: The Virtue of Forgetting in the Digital Age*, Princeton University Press.
11. Data Science Association, *Data Science Code of Professional Conduct*.
12. Chen, M. [2013], *Is 'Big Data' Actually Reinforcing Social Inequalities?*, The Nation.
13. Shin, L. [2013], *How the New Field of Data Science is Grappling With Ethics*, SmartPlanet.
14. Schutt, R. and O'Neill, C. [2013], *Doing Data Science: Straight Talk From the Front Line*, O'Reilly.
15. O'Neill, C. [2016], *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown.
16. Chang, R.M., Kauffman, R.J., Kwon, Y. [2014], *Understanding the paradigm shift to computational social science in the presence of big data*, Decision Support Systems, 63:67–80, Elsevier.

# References


17. Hurlburt, G.F., Voas, J. [2014], *Big Data, Networked Worlds*, IEEE Computer Society.
18. Introna, L.D. [2007], *Maintaining the reversibility of foldings: Making the ethics (politics) of information technology visible*, Ethics and Information Technology, 9:11–25, Springer.
19. Floridi, L. [2011], *The philosophy of information*, Oxford University Press.
20. Floridi, L. (ed) [2006], *The Cambridge handbook of information and computer ethics*, Cambridge University Press, 2006.
21. Big Data & Ethics
22. Mason, H. [2012], What is a Data Scientist?, Forbes.
23. Provost, F., Fawcett, T. [2013], *Data Science for Business*, O'Reilly.
24. Hastie, T., Tibshirani, R., and J. Friedman [2008], *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer.

# References

25. Schlimmer, J.S. [1987], *Concept Acquisition Through Representational Adjustment (Technical Report 87-19)*. Department of Information and Computer Science, UCalifornia, Irvine.
26. Iba, W., Wogulis, J., Langley, P. [1988], *Trading off Simplicity and Coverage in Incremental Concept Learning*, in Proceedings of the 5th International Conference on Machine Learning, 73-79. Ann Arbor, Michigan: Morgan Kaufmann.
27. Duch, W., Adamczak, R., Grabczewski, K. [1996], *Extraction of logical rules from training data using backpropagation networks,* in Proc. of the The 1st Online Workshop on Soft Computing, 19-30.Aug.1996, pp. 25-30.
28. Gorelik, B. [2017], Don’t study data science as a career move; you’ll waste your time!, gorelik.net.