



REAL WORLD EVIDENCE STUDY AND USE OF ANALYTICS & VISUALIZATION

Jatin Patel
Manager, Statistical Programming, PAREXEL International

PhUSE 2018

PAREXEL[®]
YOUR JOURNEY. OUR MISSION.[®]

AGENDA

REAL WORLD EVIDENCE STUDY AND USE OF ANALYTICS & VISUALIZATION

- Data Environment
- Feature Engineering
- Modeling
- Clustering
- Putting Pieces Together

The doctor of the future will give no medicine, but instead will interest his patients in the care of human frame, in diet, and in the cause and prevention of disease.

Thomas Edison (1847 – 1931)

DATA ENVIRONMENT IS RAPIDLY CHANGING

It is that we have too much data Not being used .. That's in many places .. That's hard to get to or use, When it's needed .. To do good.

Sources of the data



Mobile



EMRs



Paper / Text Documents



Social Media



Images



Videos



Sensors / Devices

40-50%

Annual growth in digital data volume*

62%

Annual growth in unstructured data*

~9X

of unstructured data vs. structured data by 2020**

+



Genome



Doctor



Pharma



Labs

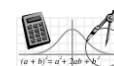
Advances analytical techniques and Machine Learning



Smarter Algorithms



Faster Processing Speed



Improved Visualization

The data from electronic health systems, social channels (unstructured and multi-dimensional data), healthcare systems in hospitals and clinics can be used to gain insights that will help better disease management plans and enhance quicker recovery. And also by combining vital signs, pathologic and genetic data, we can get the footprint for personalized medication.

* HP Autonomy, *Transitioning to a new era of human information*

** Steve Hagan, *Big data, cloud computing, spatial databases*

FEATURE ENGINEERING

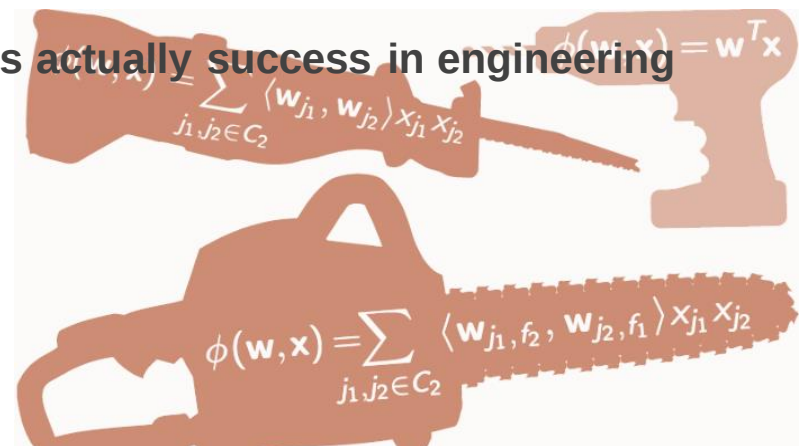
The correct use of inputs is key for a successful model.

"Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data."

- Jason Brownlee in "Discover Feature Engineering, How to Engineer Features and How to Get Good at It"

"Much of the success of machine learning is actually success in engineering features that a learner can understand."

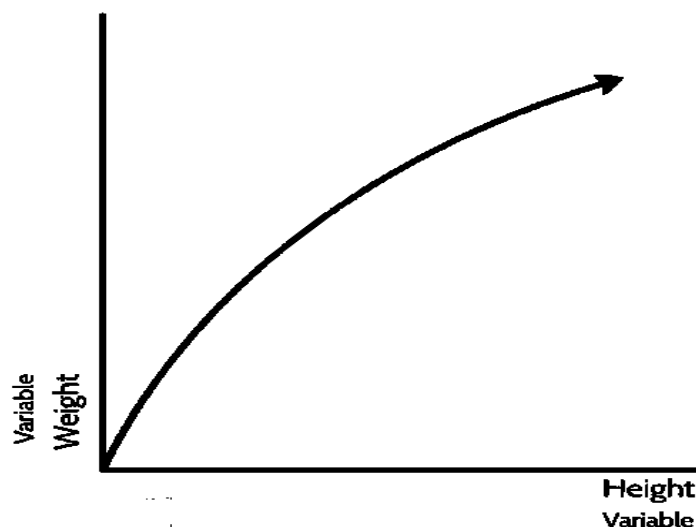
- Scott Locklin in "Neglected Machine Learning Ideas"



WHAT IS MODEL ?

Formalizes relationship between variables using mathematical equations

Taller the person is, more he/she would weight



Body mass is
proportional
Height / Square
of Height?

Remember:

Most simple
mathematical
relationships are
formal numerical
versions of our
intuitive
understanding!

Types of Models

Automated Models

Built to mechanize processes
and decisions

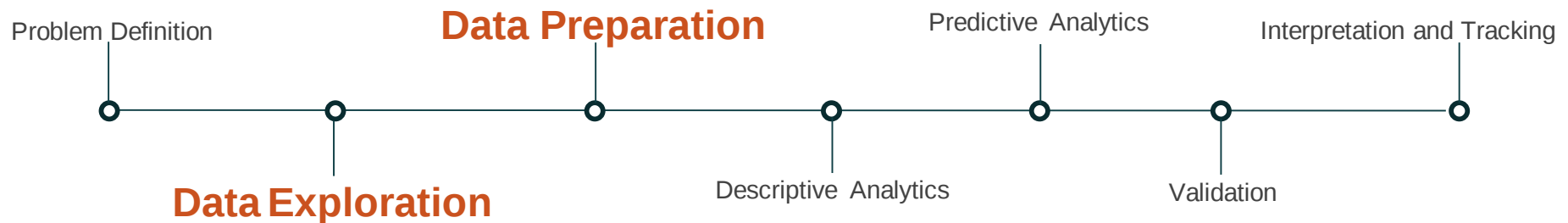
Explanatory Models

Built to make decisions,
explain decisions or justify them

CONFIDENTIAL

PAREXEL

MODELING



Data Exploration

Data Collection
Think outside of the box
Collect different types of data
Include various data sources
Important: Decide all data sources

The data dictionary,
collecting the data,
understanding the variables and quality check.

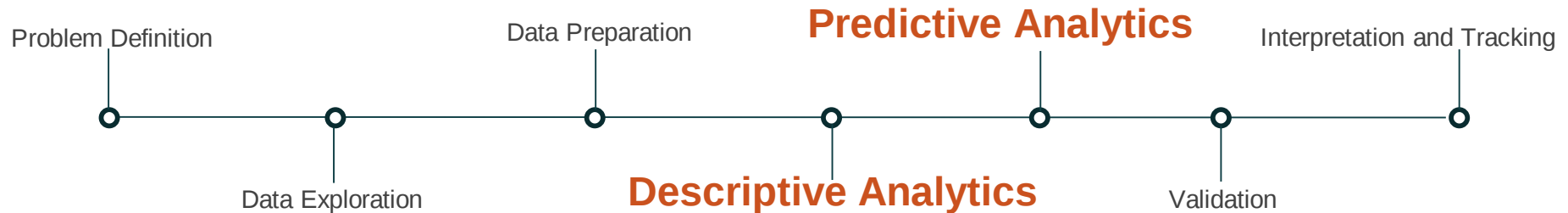
Data Preparation

Preparing the data for Analysis Think
Identify Missing Values
Treat Missing Values
Create Binary Variables
Create Other Derived Variables

Modeling techniques play a big role in influencing
the data preparation stage

Techniques like: Regression, Clustering, Decision
trees, Etc.

MODELING



Descriptive Analytics

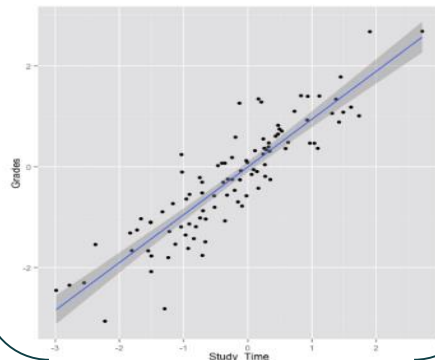
Draw Various Conclusions, Insightful Observations
Measures of
Frequency, Central Tendency, Dispersion or Variation & Position

What is Regression ?

“A Method of Measuring the link Between two or more phenomena”
In doing so. It pinpoints an “average causal effect”

Linear and Logistic Regressions

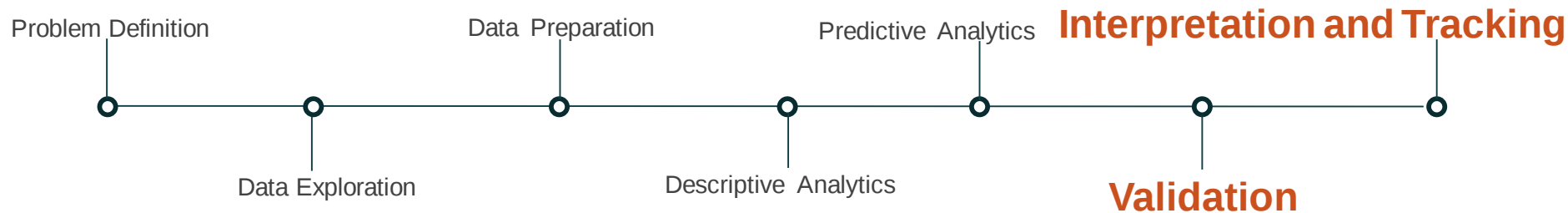
Regression: Measure link between 2 or more phenomena



Logistic regression is used to
determine
the probability of an event

	Yes	No
Heart Attack	1	0

MODELING



Model Validation

Several models can be validated to
See which one fits best!

Avoid “overfitting”

Ensures a robust and healthy
Model is selected

Work continues after the
model is built ...

Accuracy will dip over time

Model needs to be tracked and
Accordingly re-built or updated

PREDICTIVE ANALYTICS PROCESSES

Development Process

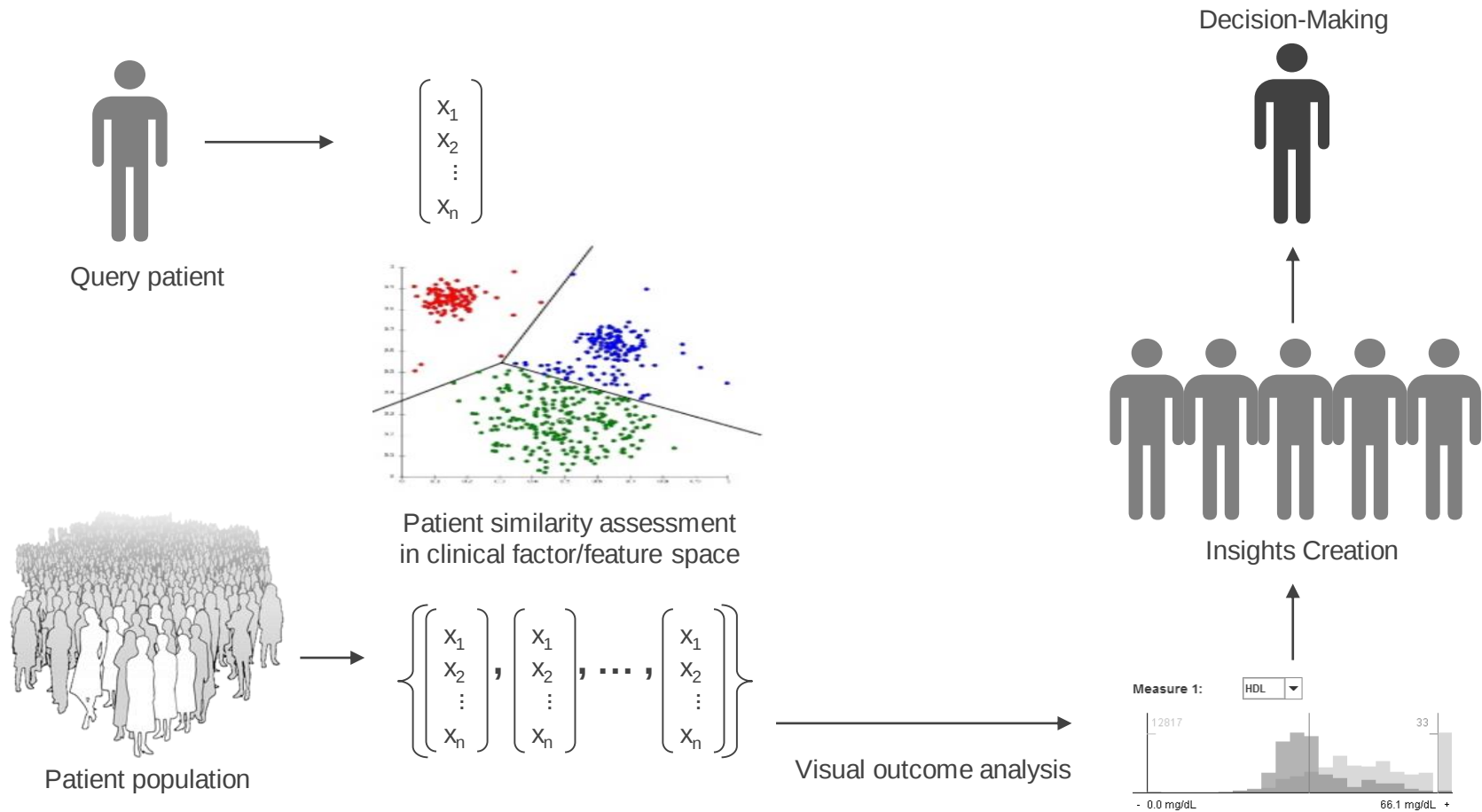


Running the Model



CLUSTERING

Clustering analysis groups objects based on their (dis)similarity and has a broad range of applications.



CONCLUSIONS

REAL-WORLD DATA IS CHALLENGING, BUT HAS GREAT POTENTIAL

- Better treatments
- More efficient care

MANY DIFFICULT CHALLENGES

- Data
- Domain-Specific Requirements

COLLABORATION ACROSS MANY DISCIPLINES

- Data analytics and statistics
- Visualization and user interaction
- Systems
- Medical domain experts

STAY 'ALIVE' AFTER DEATH

Organ donation is the process of surgically removing an organ (or tissue) from one person (the donor) and placing it into another person (the recipient). Transplantation becomes necessary when the recipient's organ fails or is damaged by a disease or some injury

 TIMES ORGAN DONATION DRIVE an initiative by THE TIMES OF INDIA	 DONOR METER 1,41,790	 Donors registered online as part of TOI's initiative
--	--	--

HELP SOMEONE TO LIVE BETTER “BEFORE” DEATH

 DATA DONATION DRIVE	
---	--

* THE TIMES OF INDIA – Times organ donation drive

THANK YOU