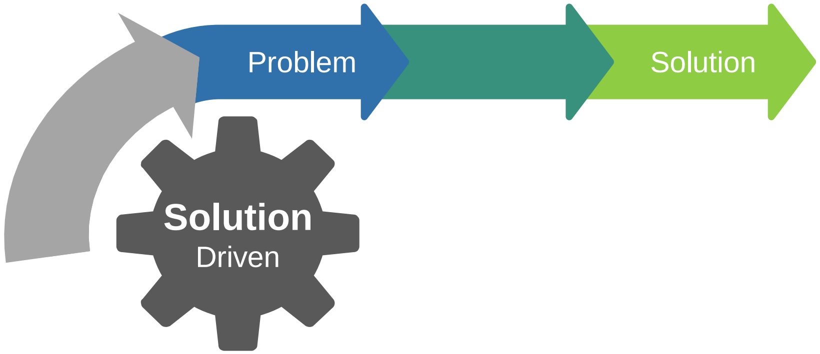
Two-Letter Acronyms With Tremendous Power: ML & AI

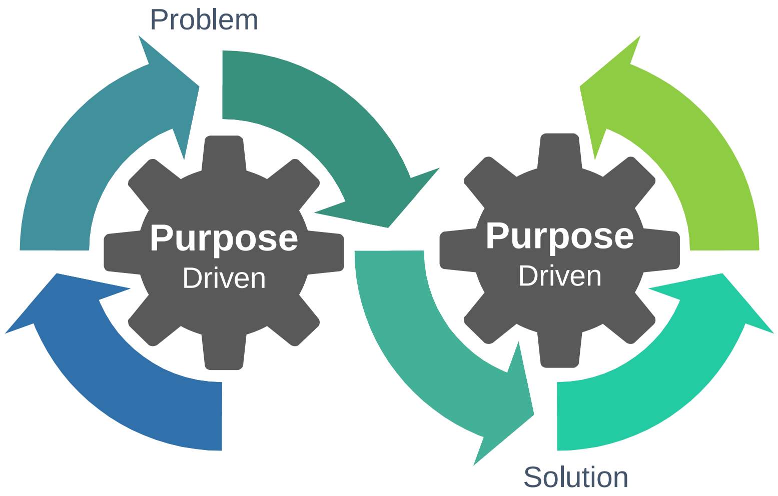
Leila Chaouki-Juneau
Ahmad Fattahi

Solution Approach

vs

Purpose Approach





Key Objectives

- ML and AI demystification
- ML and AI value added
- Methodology
- The PI System’s role
- Industry Examples

Basic Definitions

Words Soup

DO WE HAVE ANY ACTIONABLE ANALYTICS FROM OUR BIG DATA IN THE CLOUD?
Dilbert.com DilbertCartoonist@gmail.com
YES, THE DATA SHOWS THAT MY PRODUCTIVITY PLUNGES WHENEVER YOU LEARN NEW JARGON.
1-9-13 ©2013 Scott Adams, Inc./Dist. by Universal Uclick
MAYBE IN-MEMORY COMPUTING WILL ACCEL-ERATE YOUR APPLICA-TIONS.
PLUNGE, PLUNGE, PLUNGE.

What is Data Science?

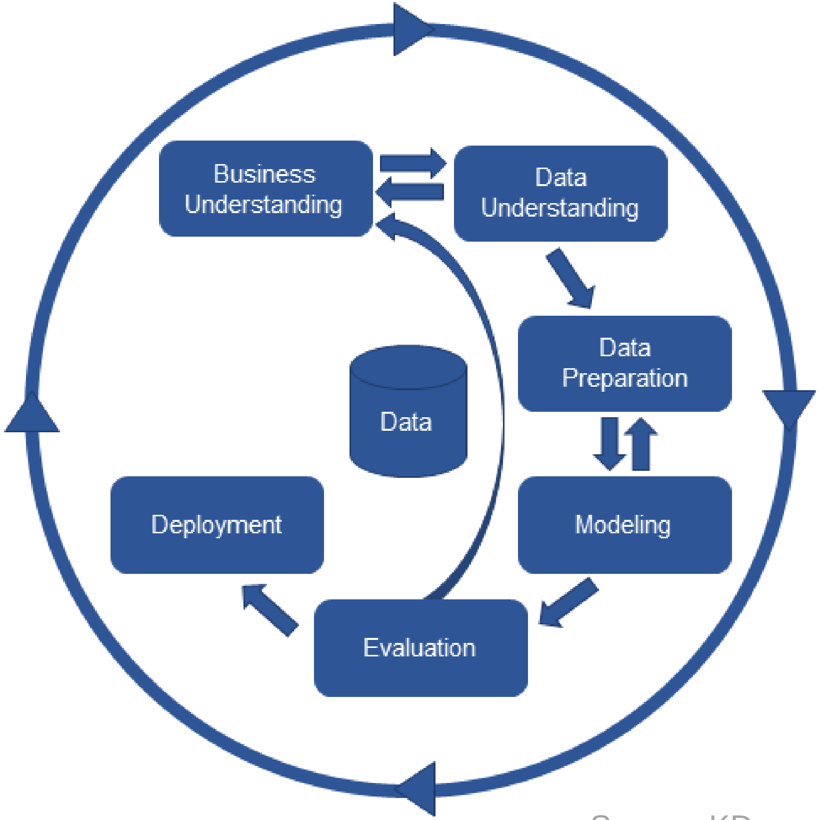
Data Science is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms.

-wikipedia

CRISP-DM

CRISP-DM

- CRoss-Industry Standard Process for Data Mining
- Imposes structure on the problem
- Emphasizes business understanding
- Highly non-linear



Source: KDnuggets

You need a Solid Business Question

- Meaningful to business leaders
- Sharp
- Supported by existing or feasible data

Different People Have Different Goals

Business Goal	Analytics Goal
Reduce the cost of production	Identify what variable contributes most to scrap material
Increase profit margin from Unit A	Optimize the choice input material based on market price/demand
Improve the quality of our hires	Identify success measures and translate into indicators on resumes
Make us safer against cyber attacks	Identify anomalies in our network traffic beyond significance threshold

“Data makes people think, emotions make them act.”

- Antonio Damasio

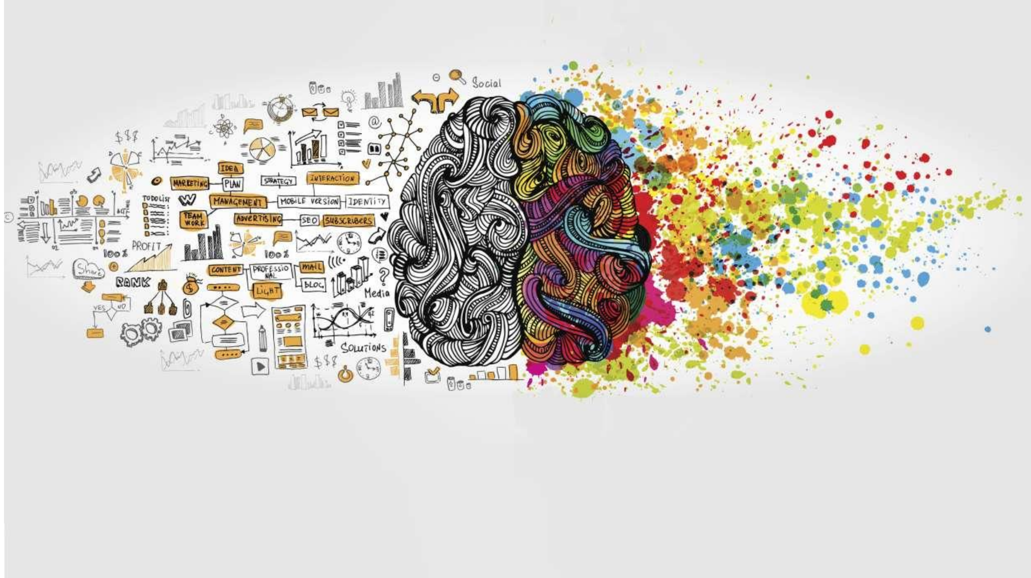
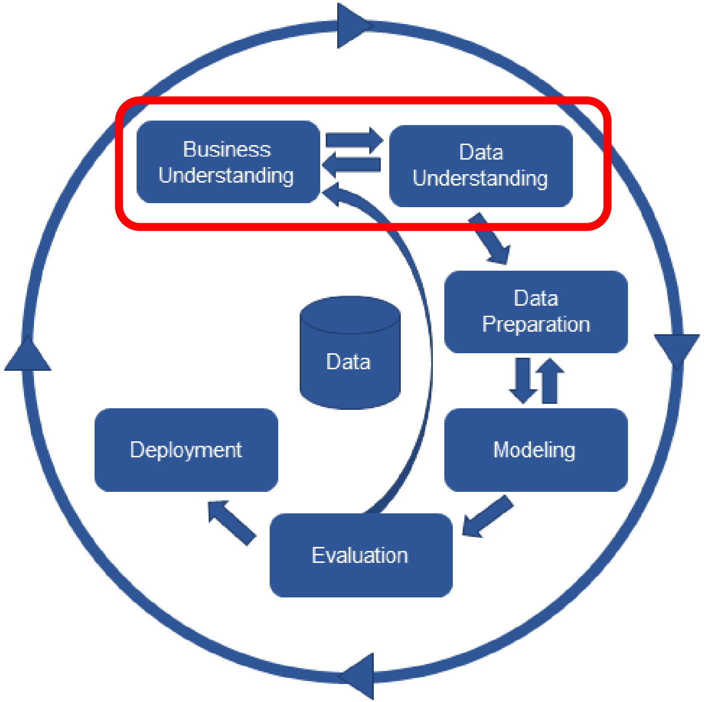


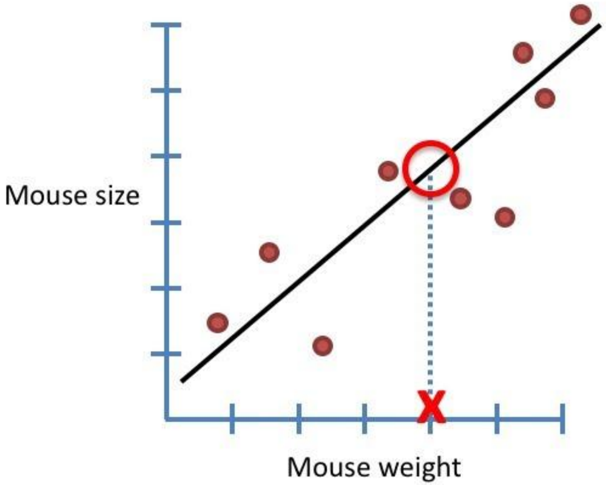
Image credit: www.mycustomer.com

Inception: Management or SME

- Start from a “Sharp Question”
- Business owner plays a key role
- Envision the delivery mechanism
- SME and data professionals start engaging
  - Many conversations until they speak the same language



What is a Statistical Model?



Building the Model

Engage with data engineers, PI Admins

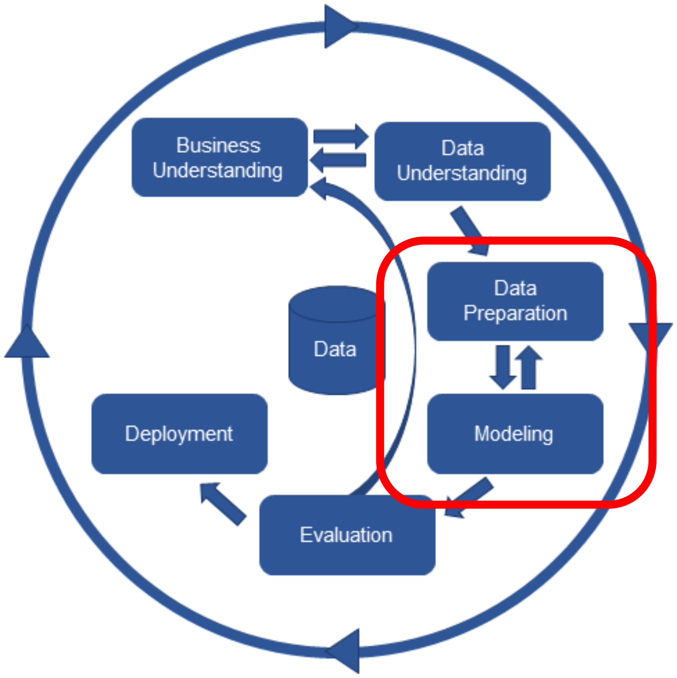
- PI Web API, AF SDK, PI Integrators, PI SQL libraries

Build the model and curate data

- Some features can be built in PI
- Build the rest in your platform

Constantly ask for validation from the SME

- Am I getting the meaning of the variables right?
- Does it make sense?



Process Data Can Be Significantly Different!

Features typically have to be engineered from raw data

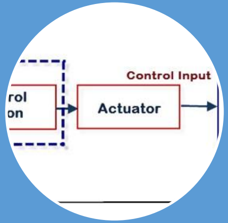
It often is not the traditional “time-series” analysis

PI System can do a lot!

- Raw, summarized, or interpolated data
- Event Frames
- Hierarchy in AF is crucial

SME plays a key role

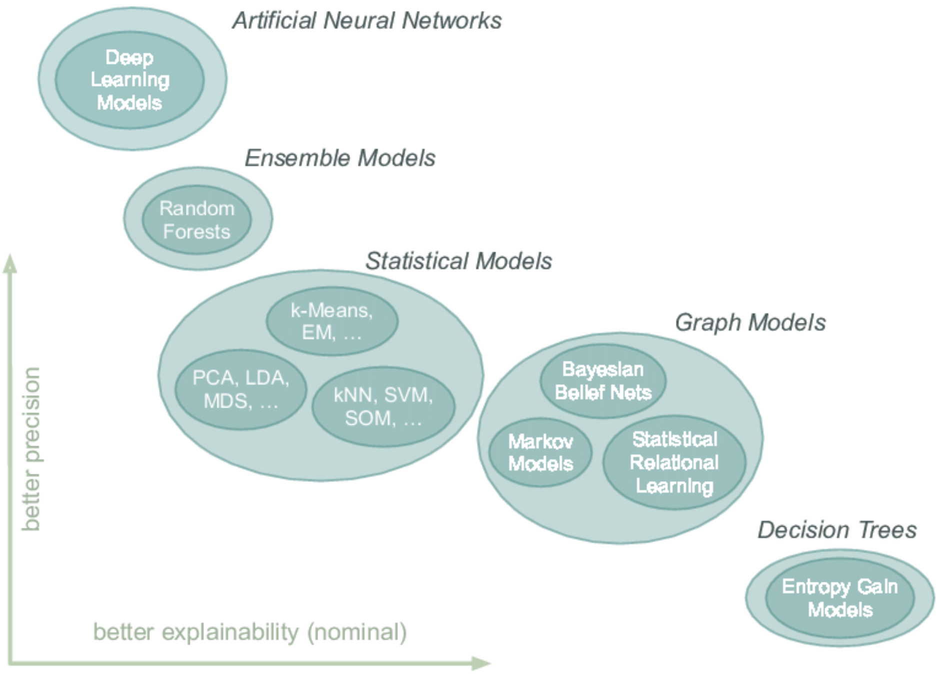
Is the goal of the project to…



control?

predict?

Tradeoff



Source: ResearchGate GmbH

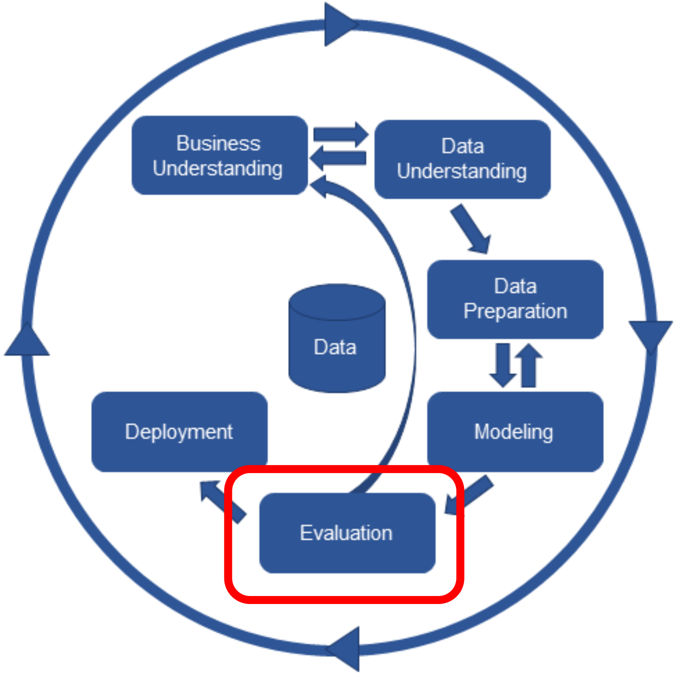
Evaluation – Loop back with the Business

Guarantees we answered the right question

Forces us to measure real value, often in dollars, man-hours, or other tangible resources

Not trivial!

Caution: data scientists speak a different language than process people



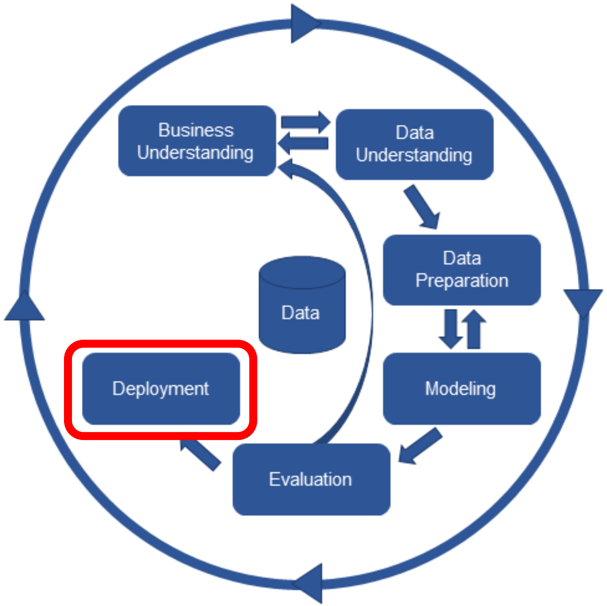
Deployment – Data Engineers Are Key

Productizing the model

Simpler models can be deployed in PI; some control models are built into the control network

Consult with PI Admins and Data Engineers early

Data Governance can pose challenges in production



Reproducible Work Is the Differentiator

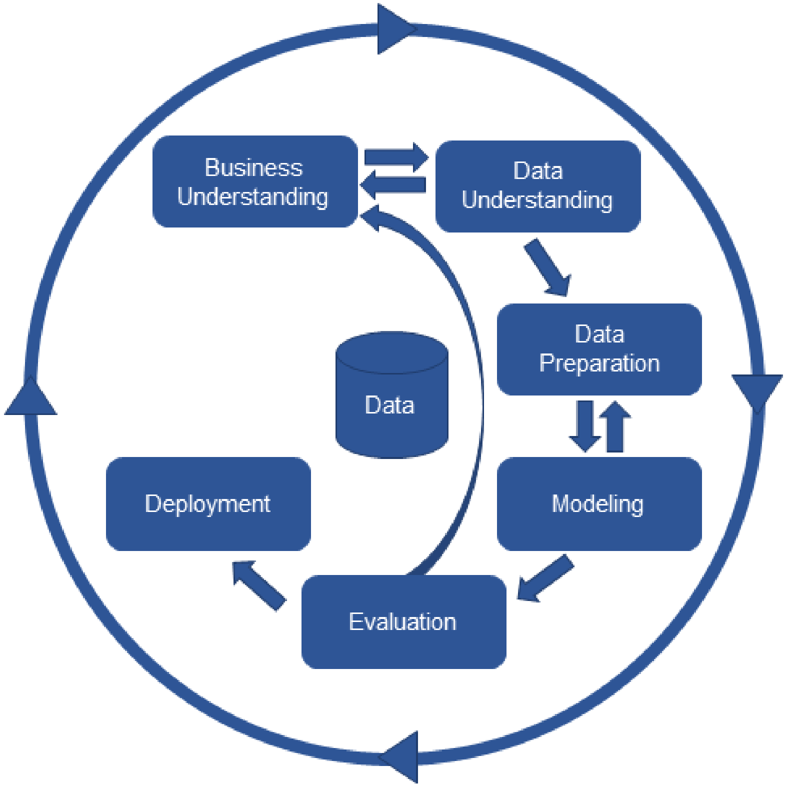
Assume your work is going to be repeated and tweaked frequently

Over time:

- Models veer off
- Physical systems change
- Priorities evolve
- New business owners come
- You get reassigned!

Leverage tools such as Jupyter Notebooks or other commercial platforms

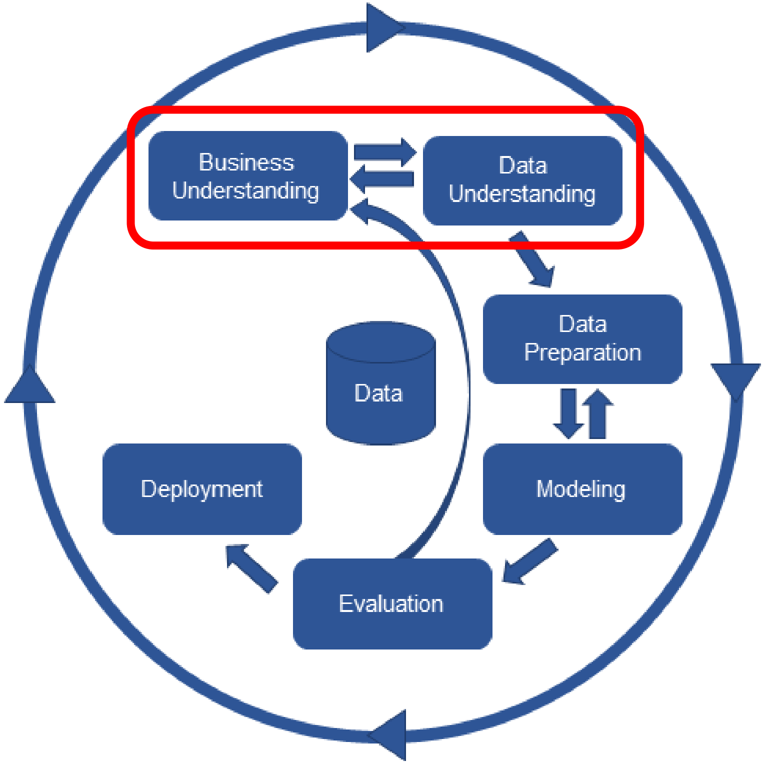
The Cycle Repeats



PI System Helps

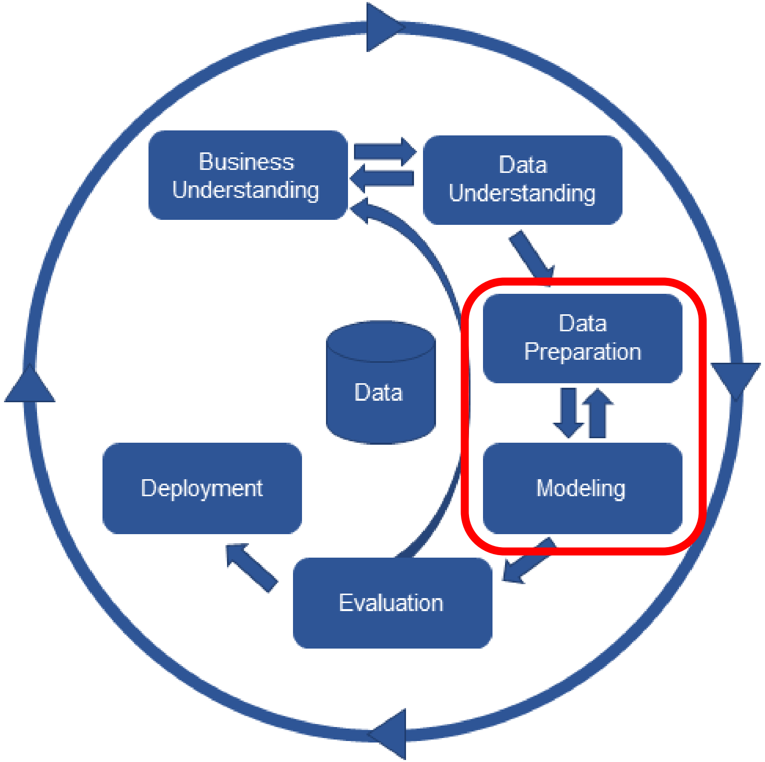
- PI Data Archive

PI System Helps



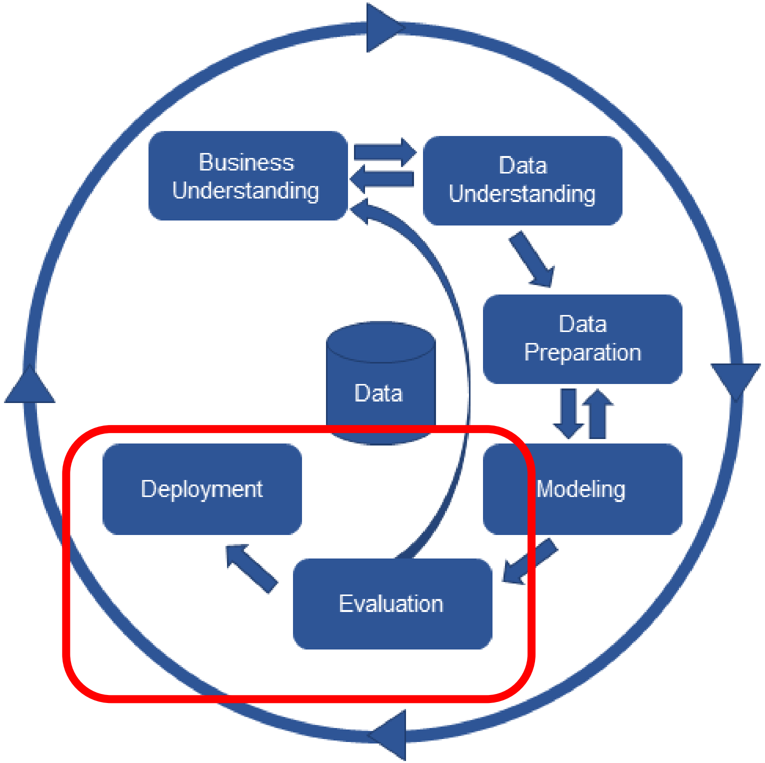
- Asset Framework
- PI Vision
- PI ProcessBook
- PI DataLink
- PI Web API
- PI SQL Client

PI System Helps



- Event frames
- PI Integrator for Business Analytics
- PI Developer Technologies

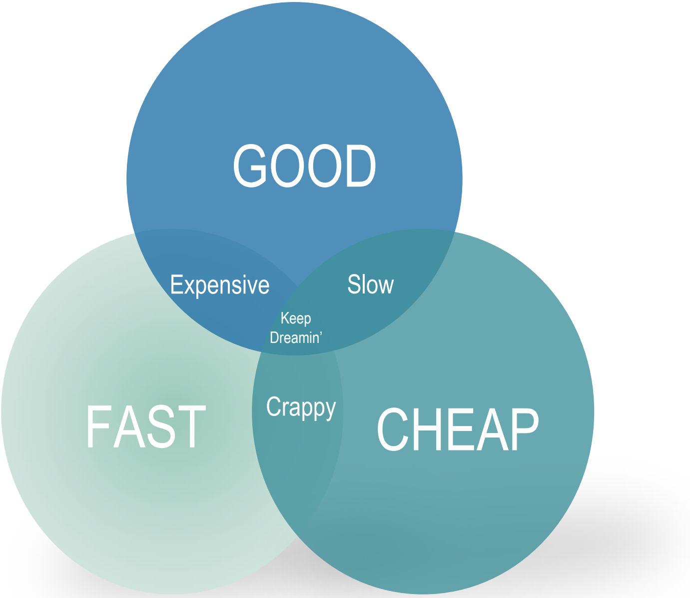
PI System Helps



- PI Integrators (Business Analytics, SAP HANA)
- PI Developer Technologies
- Asset Analytics

How do you choose a project?

- **Build or Buy** a solution
- **Technology** review
- **Intellectual Property**: Inside or Outside the organization
- **Skill** Availability: Data Science is Multidisciplinary
- **Support** of the solution
- **Cost**
- **Deployment:** Cloud or On-Premise
- **Scalability**

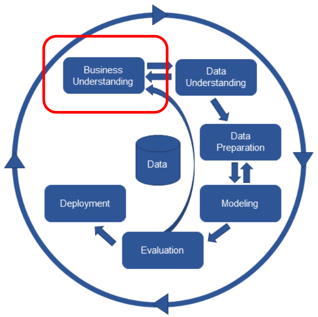


But first, the **Business Case**!

Industry Examples

Understanding Variability in Pulp Viscosity

Business Understanding



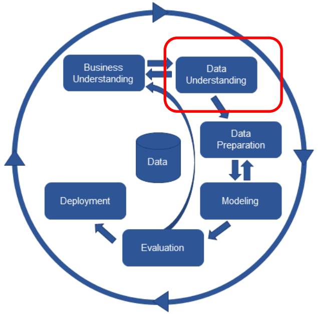
Target asset: Digester

Current situation: Visibility AFTER the fact, affecting downstream quality

Goals:
- Reduction of variability in pulp viscosity
- Deeper process understanding
- Mystical behavior discovery

Objectives:
- Build a prediction model using currant plant conditions
- Provide tools for operators to adjust cooking time accordingly
- Identify operational blind spots & inefficiencies
- Work back into upstream processes to identify & centerline key variables that directly affect digester.

Data Understanding



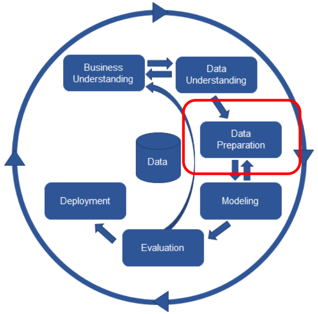
Relevance Data ➡ Brainstorm

Significant **Time Range** ➡ 1 year

Data **Sample**:

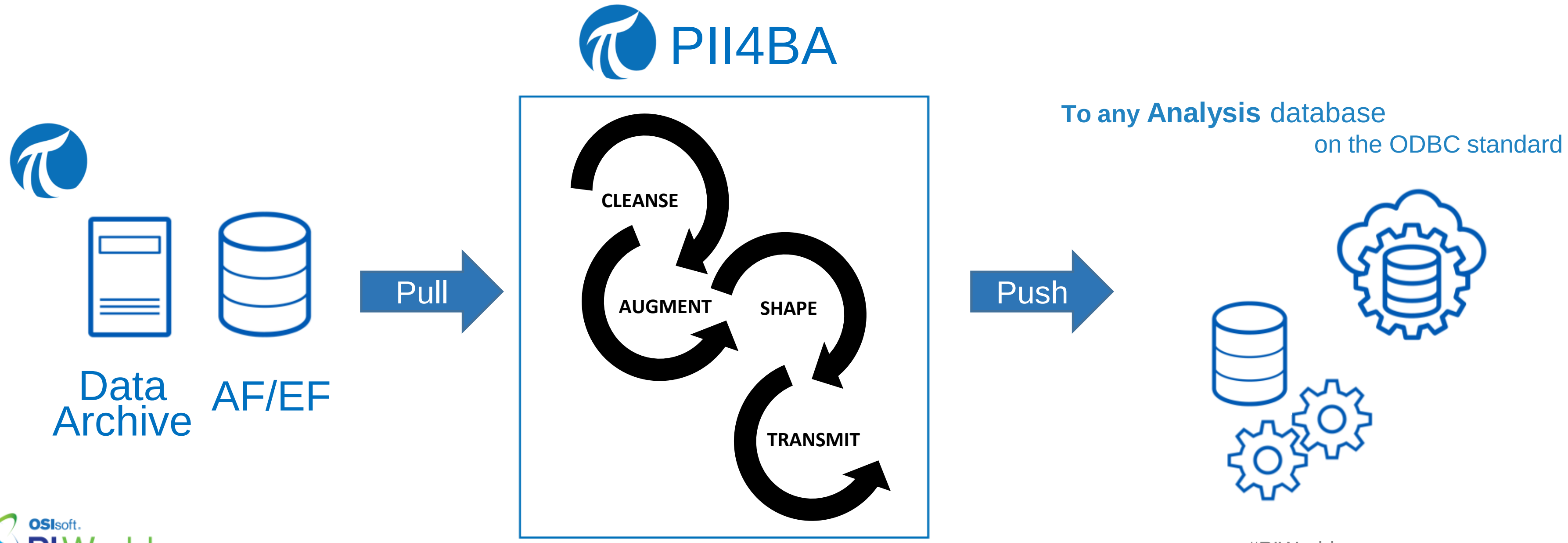
- ✓ 8000+ cooks
- ✓ 4M+ records containing 100+ variables each
- ✓ 450M+ data points

Data Preparation

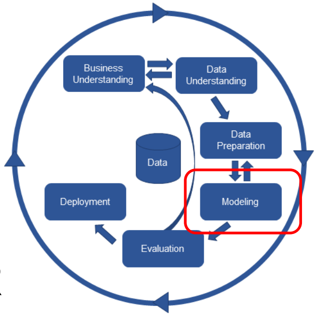


Importance of **context**:

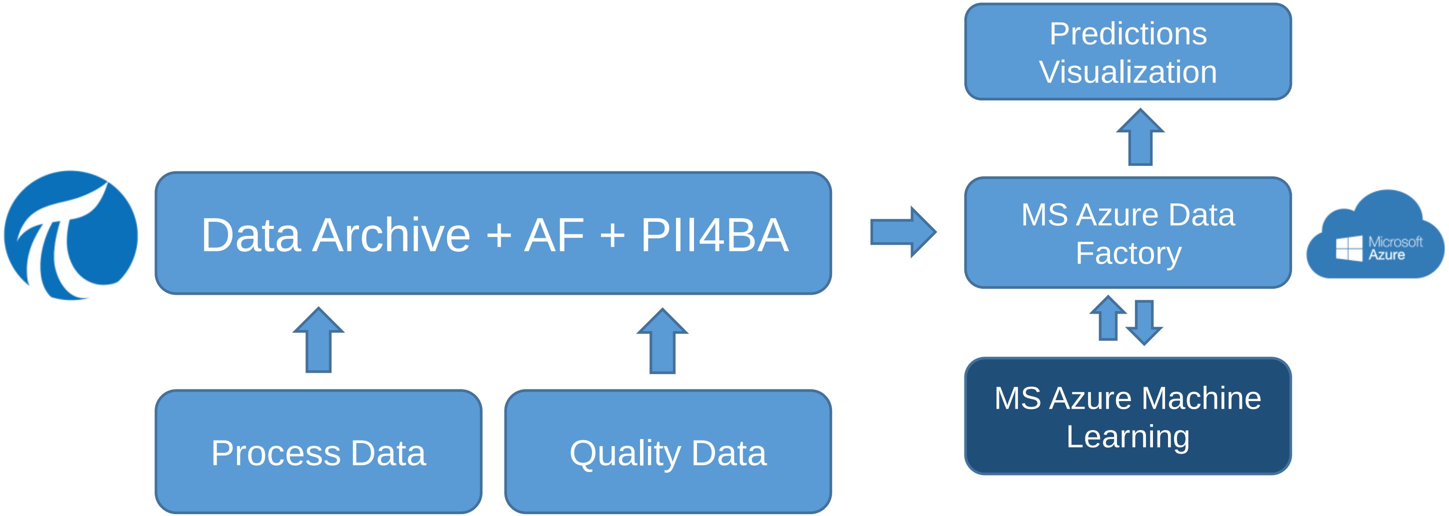
- Asset Framework (**AF**)
- Event Frames (**EF**)
- PI Integrator for Business Analytics (**PII4BA**)



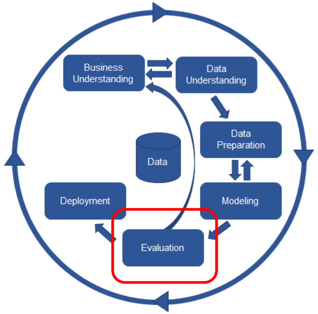
Modeling



- In house solution
- Technology: MS, Azure SQL Database, MS Azure ML, Python & R
- High Level Architecture:



Evaluation

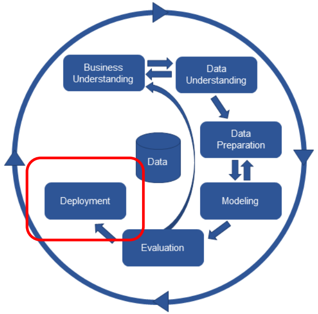


- Training the models
  - Adding additional process variables
  - Eliminating some data points

First iteration Data Sample	Second iteration Data Sample
✓ 8000+ cooks ✓ 4M+ records containing more than100 variables each ✓ 450M+ data points	✓ 20K+ cooks ✓ 10M+ records containing less than 100 variables each ✓ 100M+ data points

- Moving to live streaming to the cloud

Deployment & Key Findings



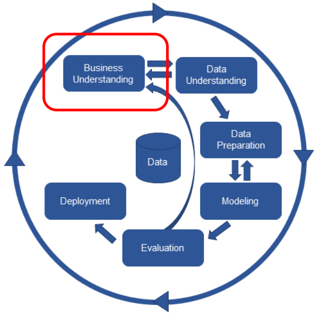
- Deployment to production with live data every 5 minutes
- Predictions only for the last cook phase
- 3 models deployed to predict viscosity
- Visualization & Reporting tools for operators to easily view the predictions and make amendments accordingly
- Ability to predict viscosity with various levels of success depending on the length of the cooks: 62%-93%

2 areas of investigation:

- **Wood Chips** loading: Adding model upstream to determine chips density, moisture and lignin S to G ratio
- **Liquor loading**: operators doing task in different order affects viscosity

Understanding Reel Quality

Business Understanding



Target asset: Paper Machine

Current situation: Variability/decrease in reel quality

Goals:
- Smoothness variability understanding
- Smoothness decrease and control

Objectives:
Control at 2 levels:
- Building Predictive model to maintain important parameters running along trajectories
- Building Optimization model to determine best trajectory for an optimal final reel quality

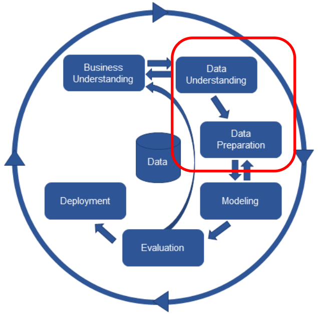
Data Understanding

Relevance Data ➡ Brainstorm

Significant **Time Range** ➡ 1 year

Frequency of data:
Process data measured in real time vs quality data quantified at the completion of a reel

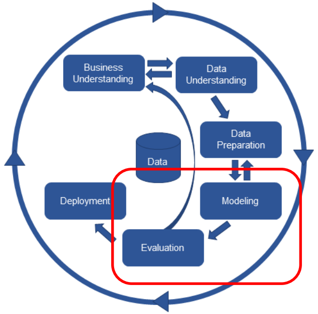
Data **Sample**:
- ✓ Single paper product
- ✓ 600+ reels
- ✓ 250+ process variables
- ✓ 5 quality variables



Data Preparation

Importance of **context**:
- Asset Framework
- Event Frames

Modeling & Evaluation



- Outsourced
- Multivariate PCA modeling

1 Correlative Variables

Predictive Model

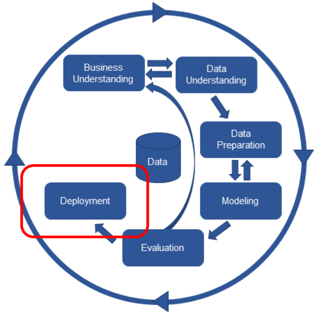
2 Causal Variables

Optimized Model

Causal helps identify manipulated variables that can really be at the core of an Optimized model to get the best outcome possible

- Bump Test used to evaluate model

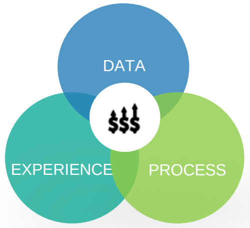
Deployment & Key Findings



- Deployment to a paper machine for performance testing
- Model predictions became part of open loop, but eventually could be in a closed loop
- Operators able to see where they are operating and where does the model suggests they run for an optimized quality
- Visualization tools give actionable recommendations if smoothness is not on target

- Quality got back to its original target
- Ability to maintain smoothness & other quality metrics
- Reduction of operational costs & more efficient use of resources

Key Takeaways

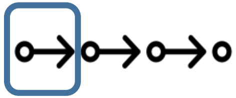


The winning combination is:
In-depth knowledge of Process
+ In-depth knowledge of the data
+ Experience

Look outside the box: Traditional process knowledge is vital but not enough

More is NOT better. Bringing key variables from the start can reduce variation

Expect Turbulence going from the Learning model to the Operational Models



Add modeling upstream for best results

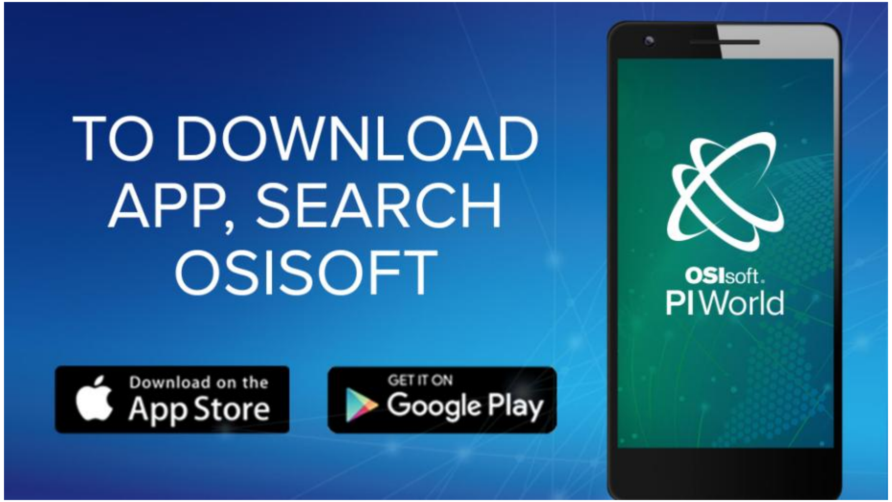
Do NOT underestimate the methodology

Questions?

Please wait for
the **microphone**

State your
name & company

Please remember



Speakers

- Ahmad Fattahi

  Manager, Data Science Enablement

  OSIsoft

  afattahi@osisoft.com

- Leïla Chaouki-Juneau

  Systems Engineer, Industry Champion P&P

  OSIsoft

  lchaouki-juneau@osisoft.com

謝謝 KEA LEBOHA
DZIĘKUJĘ CI
TAPADH LEIBH 고맙습니다
GRÀCIES
SALAMAT
БАЯРЛАЛАА MISAOTRA ANAO
NGIYABONGA
OBRIGADO شكرا
TEŞEKKÜR EDERIM
KÖSZÖNÖM
DANKIE TERIMA KASIH
DANKON TANK TAPADH LEAT
СПАСИБО
MULȚUMESC
OSIsoft PI World
FAAFETAI
РАКМЕТ СІЗГЕ
HVALA
ESKERRIK ASKO
GO RAIBH MAITH AGAT
MAHADSANID
HVALA ХВАЛА ВАМ
БЛАГОДАРЯ GRACIAS
THANK YOU
TEŞEKKÜR EDERIM
ТИ БЛАГОДАРАМ
DANK JE ΕΥΧΑΡΙΣΤΩ GRATIAS TIBI
ДЗЯКУЙ
GRAZIE
TAK DANKE
AČIŪ SALAMAT MAHALO IĀ ʻOE TAKK SKAL DU HA
DI OU MÈSI
RAHMAT MERCI
GRAZZI ÞAKKA ÞÉR
FALEMINDERIT
ありがとうございました
ĎAKUJEM
HATUR NUHUN
РАХМАТ САҒА
SIPAS JI WERE TERIMA KASIH
MATUR NUWUN
CẢM ƠN BẠN
UA TSAUG RAU KOJ
ТИ БЛАГОДАРАМ
WAZVIITA
СИПОС