Data Analytics for Artificial Lift & Production Optimization

Table of Contents

Pan American ENERGY

Data Analytics for Artificial Lift and Production Optimization

Dr. Rajan Chokshi

Dec 4-5 2023, Comodoro Rivadavia, Argentina

BAUERBERG KLEIN
E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

Safety Moment – Dooring & Dutch Reach

2

Schedule for the class – May Change!?!

	Tuesday, December 5, 2023
8:00 - 9:30	Safety Moment & Introductions 1. Digital Oil Field Data Explorations/Workflows 1.1. Digital Transformation and Oilfields 1.2. Key technologies for digital oilfields 1.3. Oilfield System Data Verification and Management 1.4. Data types in Production Domain 1.5. Data Processing Challenges 1.6. Data Basics: Cleaning, filtration, and regulation 1.7. Data Visualization 2. A Brief/Incomplete Primer on ML/AI 2.1. Data Science versus Data Analytics
9:30 - 9:50	BREAK
9:50 - 12:00	2.2. AI, ML and Deep Learning 2.3. Data Analytics Lifecycle 2.4. Bias-Variance-Complexity Tradeoff 2.5. Data Preparation 2.6. Model Types 2.7. Role of Domain Knowledge 2.8. Training & Evaluating Model 2.9 Toolsets 2.10. AI/ML Some Use Cases in Artificial Lift
12:00 - 1:00	LUNCH
1:00 - 2:30	System Setup & Checks: Google CoLab - Why do we need this? Pull datasets & codebase from Github repository 3. Rod Pump Diagnosis using Dynamometer Cards 3.1. Problem statement, Inputs, Solution Models -- SPE Paper 3.2. Hands On Exercise 4. Choke Flow Rate Study 4.1. Problem, input, and output variables definition 4.2. Hands On Exercise
2:30 - 2:50	BREAK
2:50 - 4:30	5. Flow Pattern Analysis Study 5.1. Problem, input, and output variables definition 5.2. Hands On Exercise
	Day Conclusion

	Wednesday, December 6, 2023
8:00 - 9:30	RECAP from Day 1 6. Multi-Rate Transient Analysis 6.1. Problem statement, Inputs, Solution Models 6.2. Hands On Exercise 7. Review of ESP Failure Analysis 7.1. Problem statement, Inputs, Solution Models 7.2. Solution Review 8.0. Buildup from downhole gauge data
9:30 - 9:50	BREAK
9:50 - 12:00	9. Gas Lift Optimization - Single Point Gas Lift Injection 9.1. Problem statement, Inputs, Solution Models 9.2. Hands On Exercise 10. Flow Patterns & Slug Catcher Design 10.1. Problem statement, Inputs, Solution Models 10.2. Hands On Exercise
12:00 - 1:00	LUNCH
1:00 - 2:30	11. Review of Multi-Well Optimization - SPE Paper Study 11.1. Problem statement, Inputs, Solution Models 11.2. Solution Review 12. MultiPhase Flowmeter Model 12.1. Problem statement, Inputs, Solution Models 12.2. Hands On Exercise
2:30 - 2:50	BREAK
2:50 - 4:30	13. Class Discussion - May be Exercise with your data 14. Closing Remarks
	Course Conclusion

3

Instructor intro

40 years of Global Experience

- NOC, University, Software startups & Service Companies.
- Independent consultant/advisor to Operators and Service Companies.
- Multi-phase flow, artificial lift, production surveillance & optimization.
- Three US patents and a few published papers.
- Current focus on ML/AI applications in production & facilities, Methane Emission Reduction, Workforce Competencies.

Expert Presenter/Teacher

- Twice SPE distinguished lecturer '15-'16 & '18-'19.
- Courses/seminars presented in 35+ countries for SPE, and others.
- Graduate courses taught at four US universities.

Active SPE & ALRDC volunteer

Crazy for travel: 51 countries and counting

SCAN ME

4

Now let’s learn about you…

https://ahaslides.com/PAE23

1.0. Digital Oil Field – Data Explorations and some Workflows

Dr. Rajan Chokshi

Data Analytics for Artificial Lift & Production Optimization

1

Outline

- Digital Transformation and Oilfields
- What, How, Why, When of digital oilfield projects
- Data types in Production Domain & Artificial Lift
- Data Processing Challenges
  - Data Basics: Cleaning, filtration, and regulation
- Data Visualization

2

Digital Transformation is Accelerating… Enabler Digital Oilfield…

2010 Presentations

- “up to 25% savings in OPEX, up to 8% higher production rates, 2-4% lower project costs, and as much as 6% improved resource recovery within the first full year of deployment…” [1]
- BP-Valhalla, Field of the Future®: Production increase > 1.5 MM BOE/year [2]
- Shell: “10% ***sustained*** improvement in production, 5-10% increase in recovery, 20% reduction in OPEX, and 75% reduction in workflow cycle times in core processes.” [3]

Sources:
1. IHS-CERA, Chevron’s Next magazine, Issue 5: http://bit.ly/29rFpgR
2. BP’s Field of the Future, http://on.bp.com/29EwtHp
3. Deloitte, Unearthing the potential of digital oilfield technology, http://bit.ly/29waG44

JPT JOURNAL OF PETROLEUM TECHNOLOGY

Digital Transformation at BP is Starting to Add Up to Billions

Trent Jacobs, JPT Digital Editor | 16 May 2019

JPT JOURNAL OF PETROLEUM TECHNOLOGY

How Southeast Asian Upstream Operators Are Digitalizing

Matt Zborowski, Technology Editor | 01 June 2019

JPT

AI/MACHINE LEARNING

AI Is Here, and It’s Helping With Predictive Maintenance in the Oil Field

Oil and gas operators such as Shell and Oxy are now employing AI together with a vast network of sensors and other machine-learning software to stamp out problems before they happen.

May 24, 2023 By Blake Wright
Data Science and Digital Engineering

3

3

Problems facing operators: Day-to-day operational questions

- Are my wells / equipment delivering per plan?
  - Any hurdles/problems I need to prepare and plan for?
  - Any impending failures so that I can plan?
  - Which lease operating costs are eating my profitability?
- Production forecast? Are we meeting our production quotas?
- Which installations and activities are causing production deferrals?
- Safety of my people? Environmental issues???
- Can I put in place all the best practices on a sustaining basis???

Source: Edwards T, SPE Webinar 09/2017, DO in Business, https://tinyurl.com/y4amup8z

4

SPE International Society of Petroleum Engineers

SPE ATCE SINCE 1924 | 16–18 October 2023 San Antonio, Texas, USA

1.1. What, Why, When, How of the digital oilfield projects

Accutant Solutions
Accurate Accountable Acumen

5

What is Digital Oilfield (DOF)?

- 'The vision for the Digital Oil Field is one where operators, partners, and service companies seek to take advantage of improved data and knowledge management, enhanced analytical tools, real-time systems, and more efficient business processes'

IHS CERA: Digital Oil Field of the Future

- What is enabled by DOF?
  - Moving to a real time or near real time way of working
  - Connection of one or more remote sites or teams to work together
  - Moving to more multidiscipline way or working
  - Value chain integration and optimization

'How we run our companies in the future'

Source: Edwards T, SPE Webinar, DO in Business, https://tinyurl.com/y4amup8z

6

DOF for Artificial Lift & Production

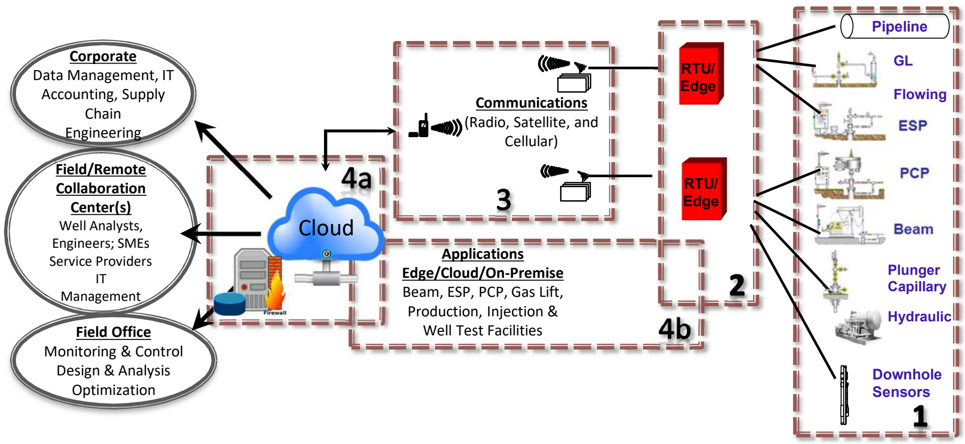


Image Source: Weatherford

7

Overall Framework

8

1. **SENSE:** Instrumentation + automation at the well head & field facilities
2. **COMMUNICATE:** Remote monitoring and control of field instruments
3. **STORE:** Localized and remote/cloud-based data stores & access
4. **ANALYZE:**
   - Data Analytics@ EDGE/ Cloud / On Premise
   - Data preprocessing: Validation & cleaning; Machine learning
   - Modeling of wells, subsurface, surface network, production plant
     - Optimization software for artificial lift solutions
     - Integration with other systems for reservoir, completion, well test, drilling, etc.
     - Data driven approaches
   - Visualizations, Dashboards
5. **DECIDE:** Collaborative Environments

This framework requires full IT & OT (Operations Technology) support

8

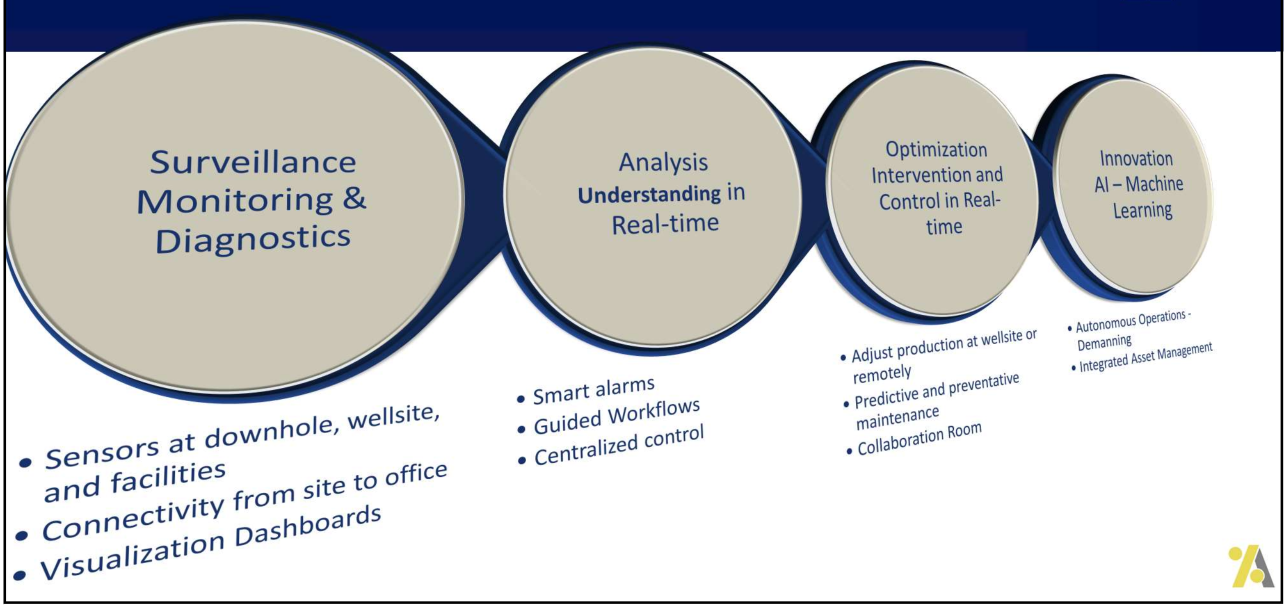
Opportunities to improve Production Operations using Data

Best Practice	Functional Capabilities	Benefits/Business Case
Exception-Based Predictive Surveillance	• Monitor volumes and rates by well, reservoir, facilities, route, lift type, etc. by shift/daily/weekly/monthly/quarterly • Trend Production, target and deviation with event history • Generate multilevel notifications based on operations-driven thresholds	• Manage well proactively for faster decisions • Identify problem wells in real time • Monitor deviations from targets • Analyze production issues in context • Diagnose operating issues • Improve communications
Well Target Setting	• Visualization of production history and trends • Identification of production cycles based on surface constraints • Modified decline curve analyses • Monitor deviations, target history and field operations to identify revised targets	• More realistic target setting based on access to well production history and operating constraints • Targets refreshed per business requirements
Tracking and Optimize Uplift	• Track jobs related to well production enhancement (e.g. workover) • Monitor before/after production (and forecast) • Access comparative effectiveness by job/well	• Evaluation and Comparison of workover benefit by well, type of job, type of well, well grouping • Cost-benefit economic analysis
Facilities Management	• Define operating envelope for separators, tanks, compressors, pumps, etc • Establish thresholds based on changing operating conditions • Dynamically determine time to capacity (or other limits) based on real-time operations; Generate operator notification	• Minimize curtailment/shut-in due to unavailability of equipment and/or facilities • Improve QHSE

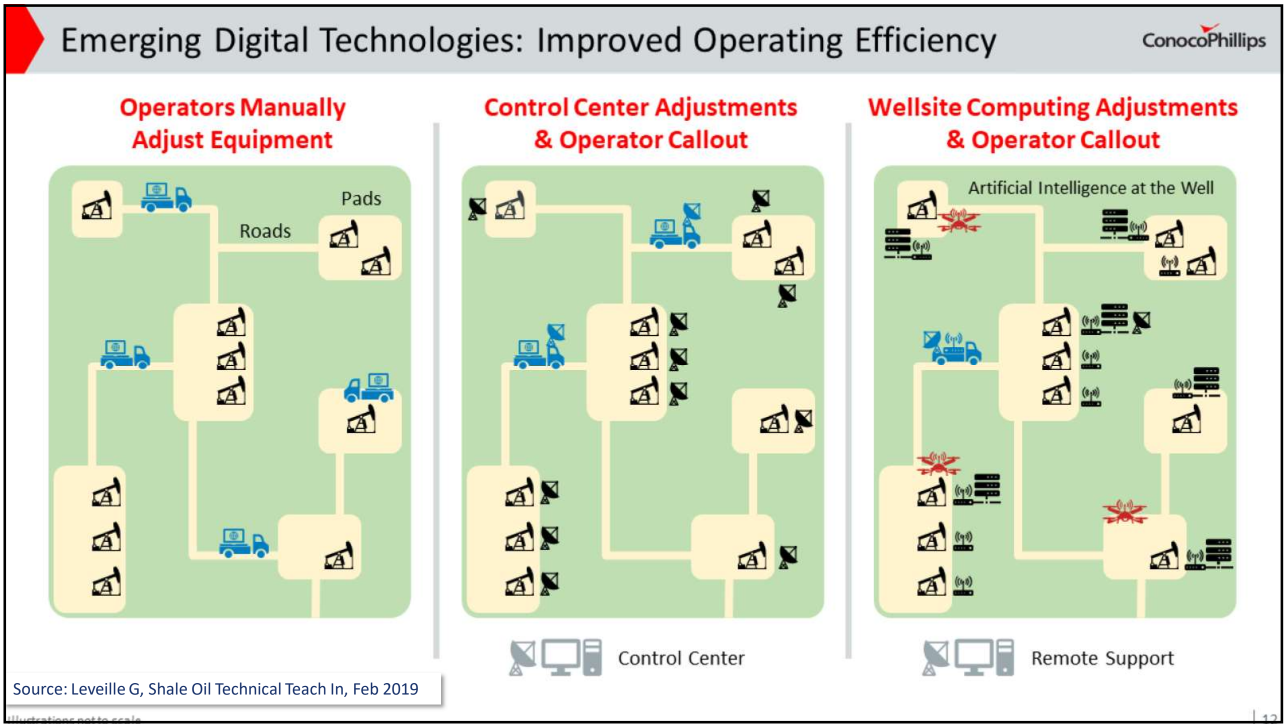
Source: Chris Lenzsch, EMC^2 http://bit.ly/1Onql7P

9

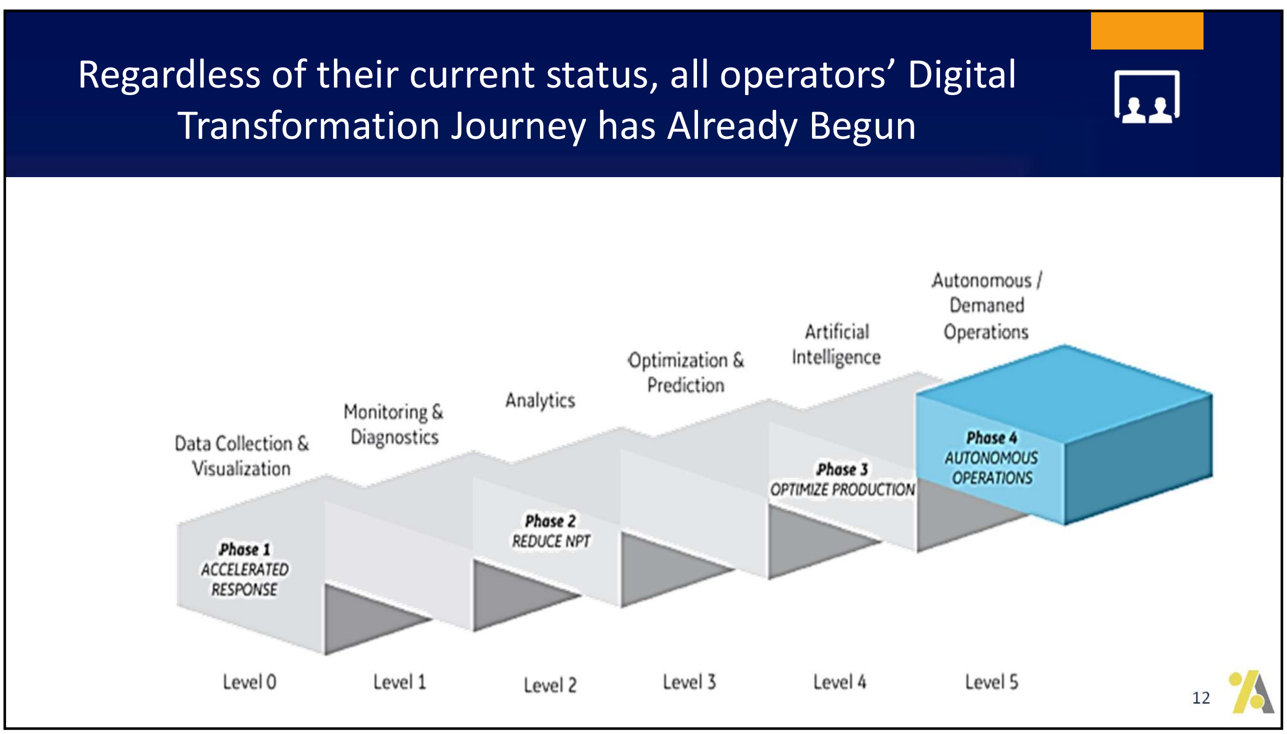
Digital Oil Field Progression



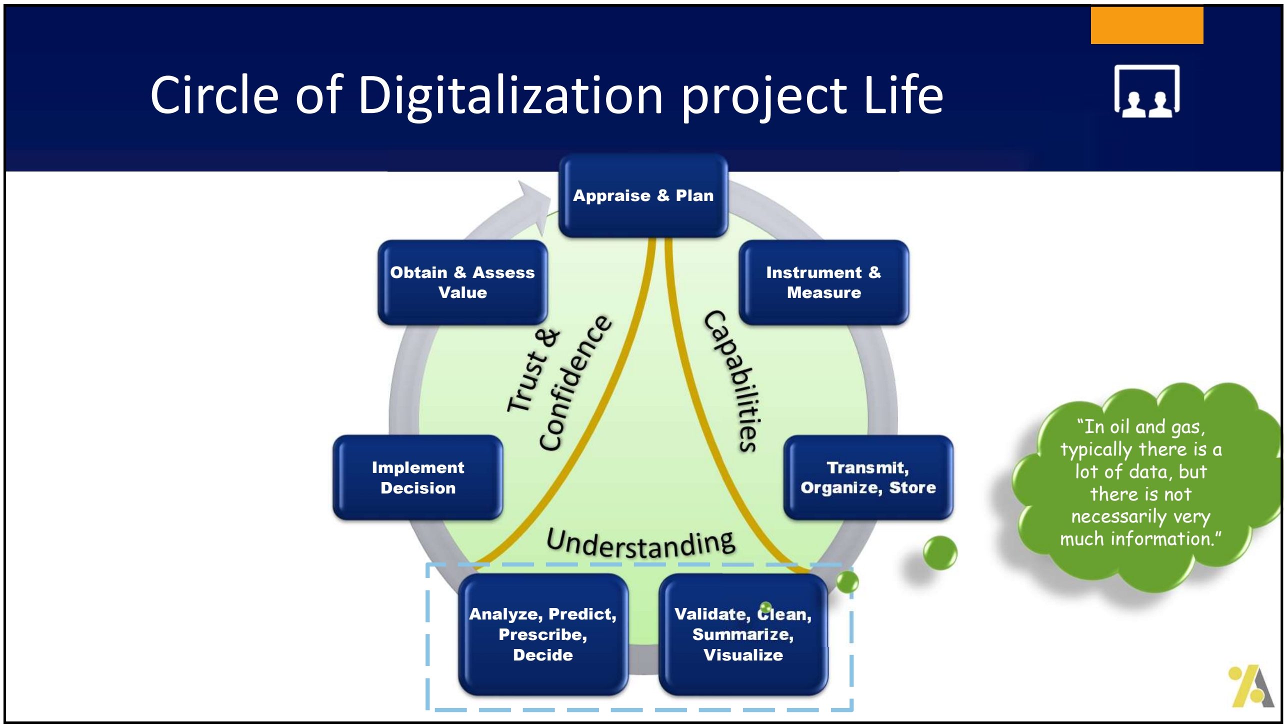
10


Emerging Digital Technologies: Improved Operating Efficiency
ConocoPhillips
Operators Manually Adjust Equipment
Control Center Adjustments & Operator Callout
Wellsite Computing Adjustments & Operator Callout
Pads
Roads
Artificial Intelligence at the Well
Control Center
Remote Support
Source: Leveille G, Shale Oil Technical Teach In, Feb 2019

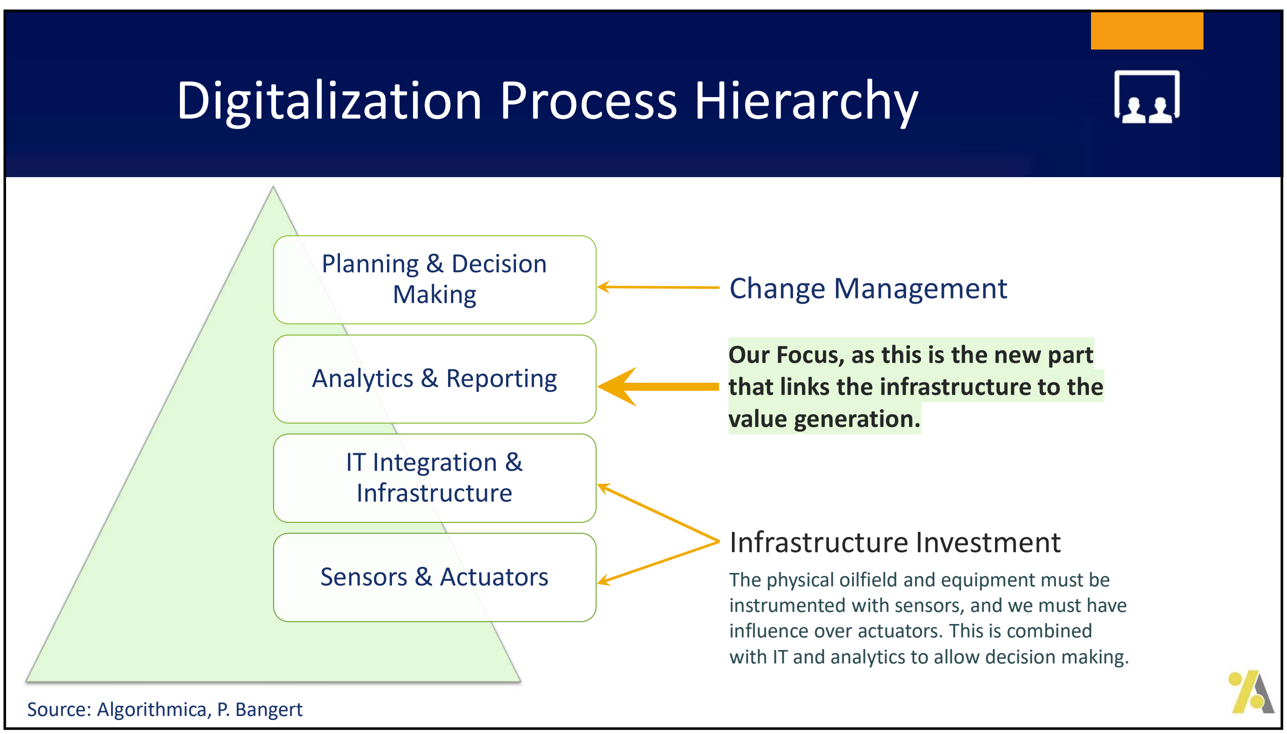
11


Regardless of their current status, all operators' Digital Transformation Journey has Already Begun
Data Collection & Visualization
Monitoring & Diagnostics
Analytics
Optimization & Prediction
Artificial Intelligence
Autonomous / Demaned Operations
Phase 1 ACCELERATED RESPONSE
Phase 2 REDUCE NPT
Phase 3 OPTIMIZE PRODUCTION
Phase 4 AUTONOMOUS OPERATIONS
Level 0
Level 1
Level 2
Level 3
Level 4
Level 5
12

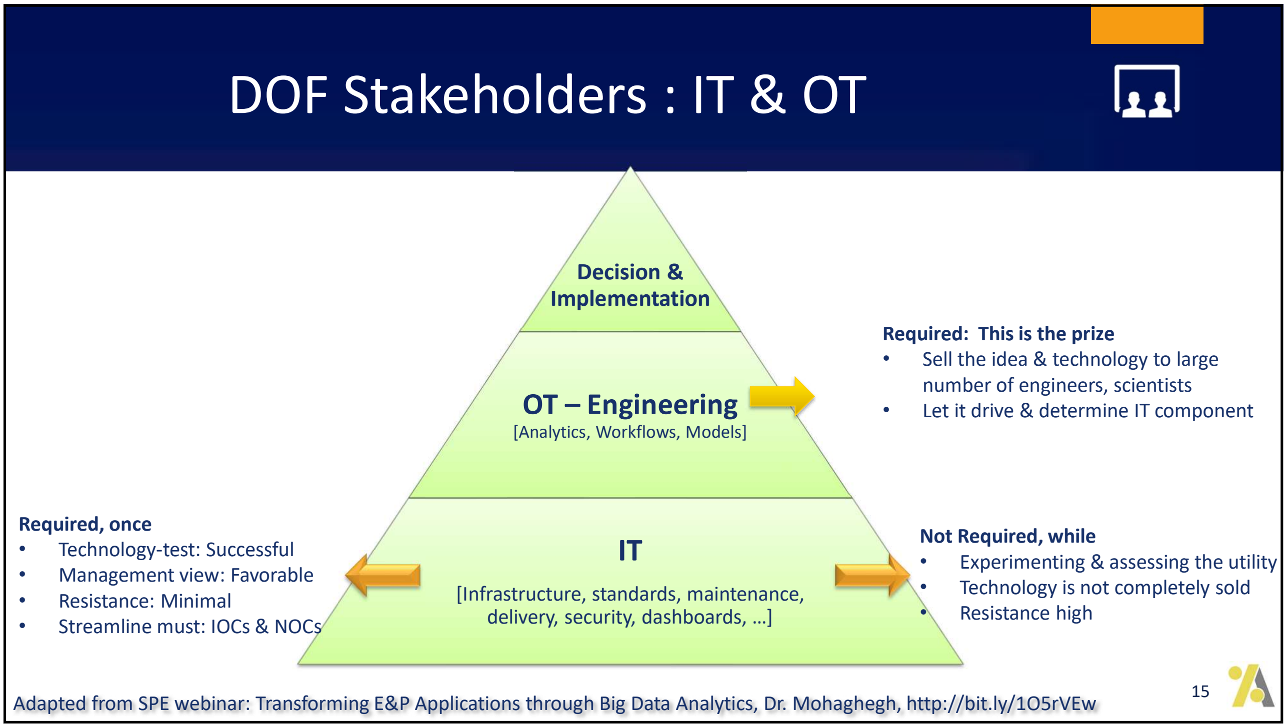
12


Circle of Digitalization project Life
Appraise & Plan
Instrument & Measure
Transmit, Organize, Store
Validate, Clean, Summarize, Visualize
Analyze, Predict, Prescribe, Decide
Implement Decision
Obtain & Assess Value
Trust & Confidence
Capabilities
Understanding
"In oil and gas, typically there is a lot of data, but there is not necessarily very much information."

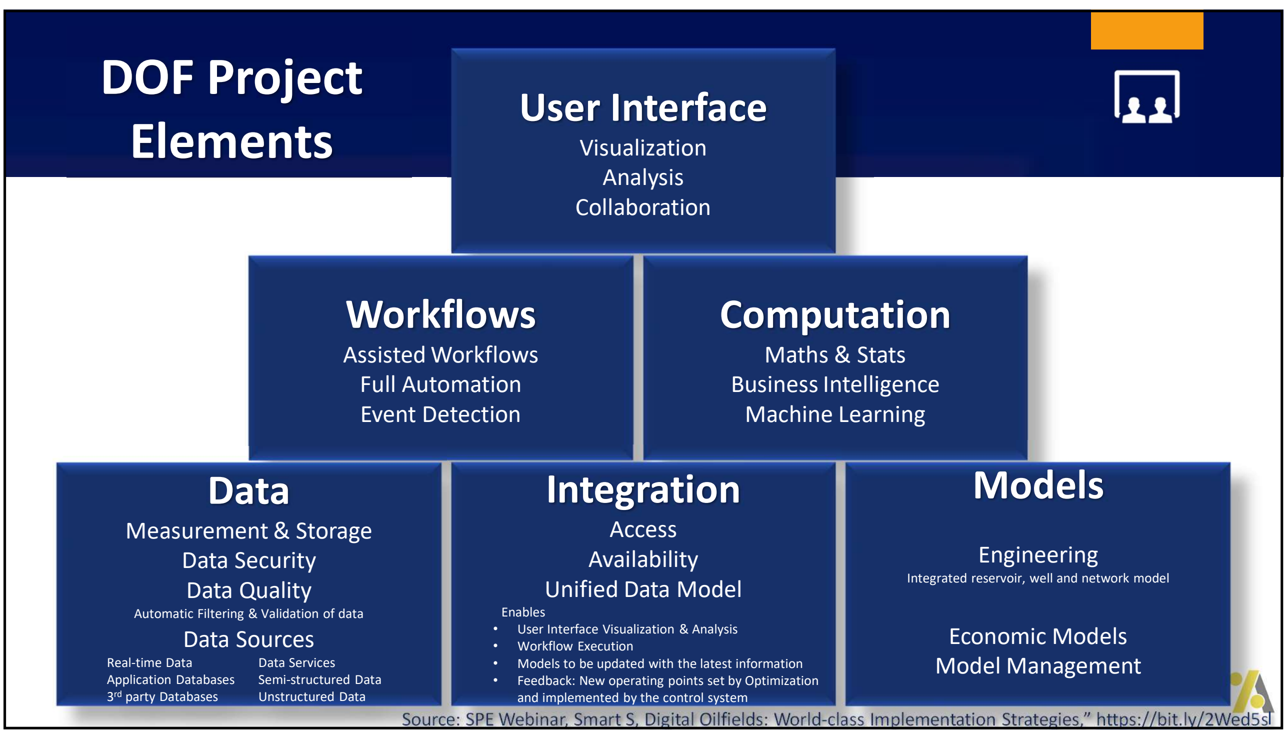
13


Digitalization Process Hierarchy
Planning & Decision Making
Analytics & Reporting
IT Integration & Infrastructure
Sensors & Actuators
Change Management
Our Focus, as this is the new part that links the infrastructure to the value generation.
Infrastructure Investment
The physical oilfield and equipment must be instrumented with sensors, and we must have influence over actuators. This is combined with IT and analytics to allow decision making.
Source: Algorithmica, P. Bangert

14


DOF Stakeholders : IT & OT
Decision & Implementation
OT – Engineering
[Analytics, Workflows, Models]
IT
[Infrastructure, standards, maintenance, delivery, security, dashboards, …]
Required: This is the prize
• Sell the idea & technology to large number of engineers, scientists
• Let it drive & determine IT component
Required, once
• Technology-test: Successful
• Management view: Favorable
• Resistance: Minimal
• Streamline must: IOCs & NOCs
Not Required, while
• Experimenting & assessing the utility
• Technology is not completely sold
• Resistance high
Adapted from SPE webinar: Transforming E&P Applications through Big Data Analytics, Dr. Mohaghegh, http://bit.ly/1O5rVEw
15

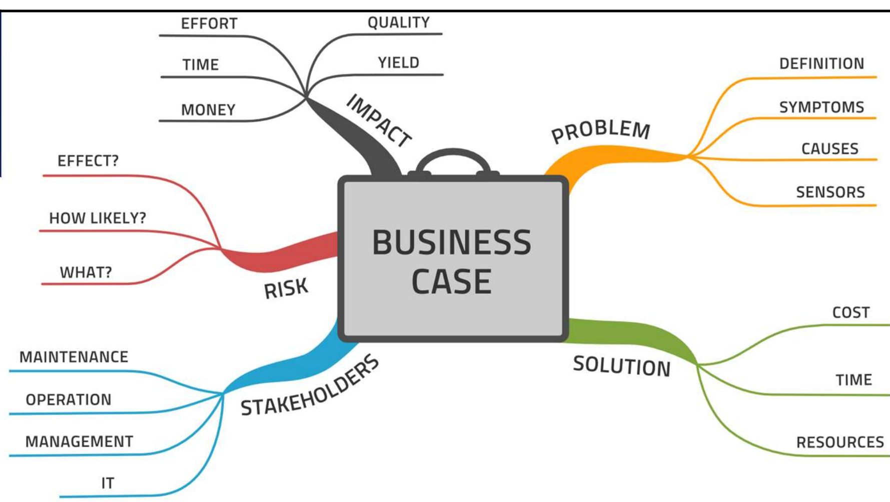
15


DOF Project Elements
User Interface
Visualization
Analysis
Collaboration
Workflows
Assisted Workflows
Full Automation
Event Detection
Computation
Maths & Stats
Business Intelligence
Machine Learning
Data
Measurement & Storage
Data Security
Data Quality
Automatic Filtering & Validation of data
Data Sources
Real-time Data
Application Databases
3rd party Databases
Data Services
Semi-structured Data
Unstructured Data
Integration
Access
Availability
Unified Data Model
Enables
• User Interface Visualization & Analysis
• Workflow Execution
• Models to be updated with the latest information
• Feedback: New operating points set by Optimization and implemented by the control system
Models
Engineering
Integrated reservoir, well and network model
Economic Models
Model Management
Source: SPE Webinar, Smart S, Digital Oilfields: World-class Implementation Strategies," https://bit.ly/2Wed5sl

16

Do first things first….
Formulate Business Case

17



If we had a model, what impact would it have? How accurate does it need to be for that impact? Who needs to be convinced for it to be used? What are the risks? What are the costs and resource needs?

18

Time
- Faster Process
- Faster Reaction
- Earlier Detection

Effort
- Less work
- Less tedious work
- Less danger
- Less HSE impact

Quality
- Higher Process Stability
- Product specifications always met but not overdone

Yield
- Higher Yield
- Higher Selectivity
- Lower Scrap or Out-Of-Spec Product
- Less Losses or Spills

Money
- More Revenue
- Less Cost
- Market and Competitive Advantage
- Result of other dimensions

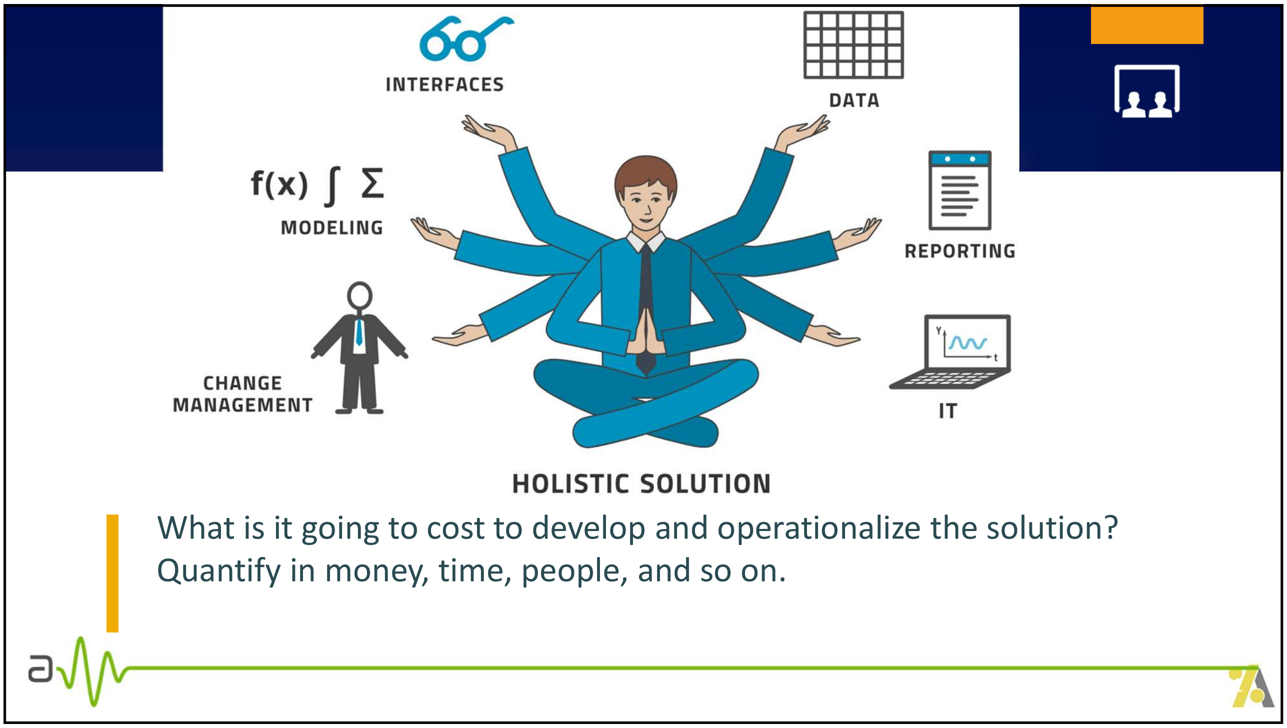
If we had the solution, what added-value impact would it deliver? Quantify this in time, effort, quality, yield, and so on to get a final monetary gain.

19

	Budget Approval	Decision / Enabler	Participant	Obstacle
Management				
Operations				
Maintenance				
Engineering				
IT				

In order to develop the solution and to get it operational, which people need to decide, enable, use, participate in the project and product? What is their interest in or against this?

20


INTERFACES
DATA
f(x) ∫ Σ
MODELING
REPORTING
CHANGE MANAGEMENT
IT
HOLISTIC SOLUTION
What is it going to cost to develop and operationalize the solution? Quantify in money, time, people, and so on.

21


Data Quality / Availability
Stakeholder Disapproval
Moving Targets
Shifting Landscape
Project Delays
What can go wrong in development and production? How likely are those factors? What impact will they have on the stakeholders, the added-value, the project/product, and the resources needed?

22

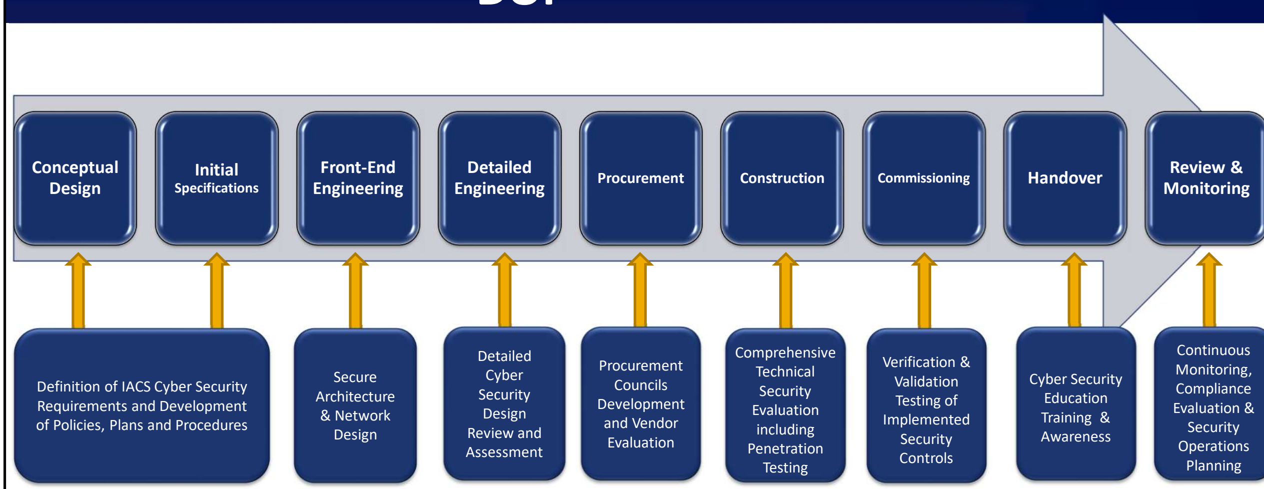
Important: Cyber Security Threat Considerations

- It goes beyond Virus Spread, Data and/or Computer Damage, or even Data Stealing.
- Think of
  - Hijacking device or entire field infrastructure → Loss of Control
  - Reporting and/or registering fake information for custody transfer → Loss of Revenues from Production
  - Unsafe field device setting → Injury or fatality or environmental incident or damage to critical infrastructure
    - Increasing pressures in pipeline or wellbores
    - Out of range speeds for rotating or moving machinery
    - Untimely opening/closing a motorized valve

Source: SPE Webinar, Al Issa I, "Protecting the Digital Oil Field from the Emerging Cyber Threats," http://bit.ly/28Tj4wJ

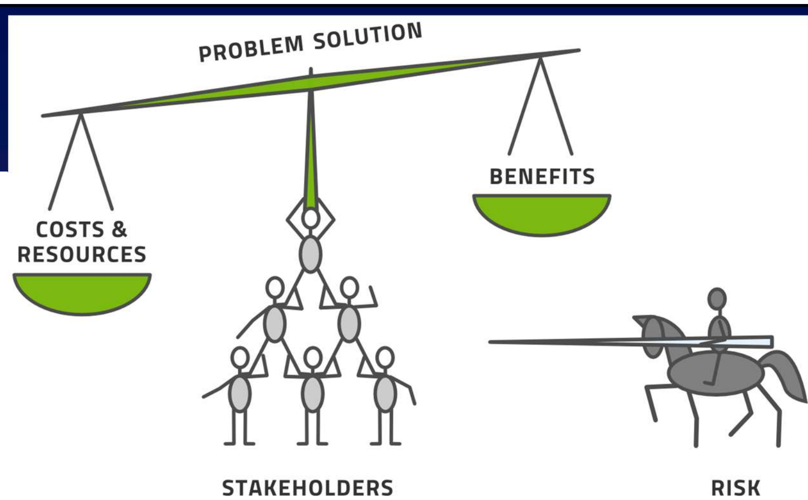
23

Must build Security into the full lifecycle of DOF



Source: SPE Webinar, Al Issa I, "Protecting the Digital Oil Field from the Emerging Cyber Threats," http://bit.ly/28Tj4wJ 24

24



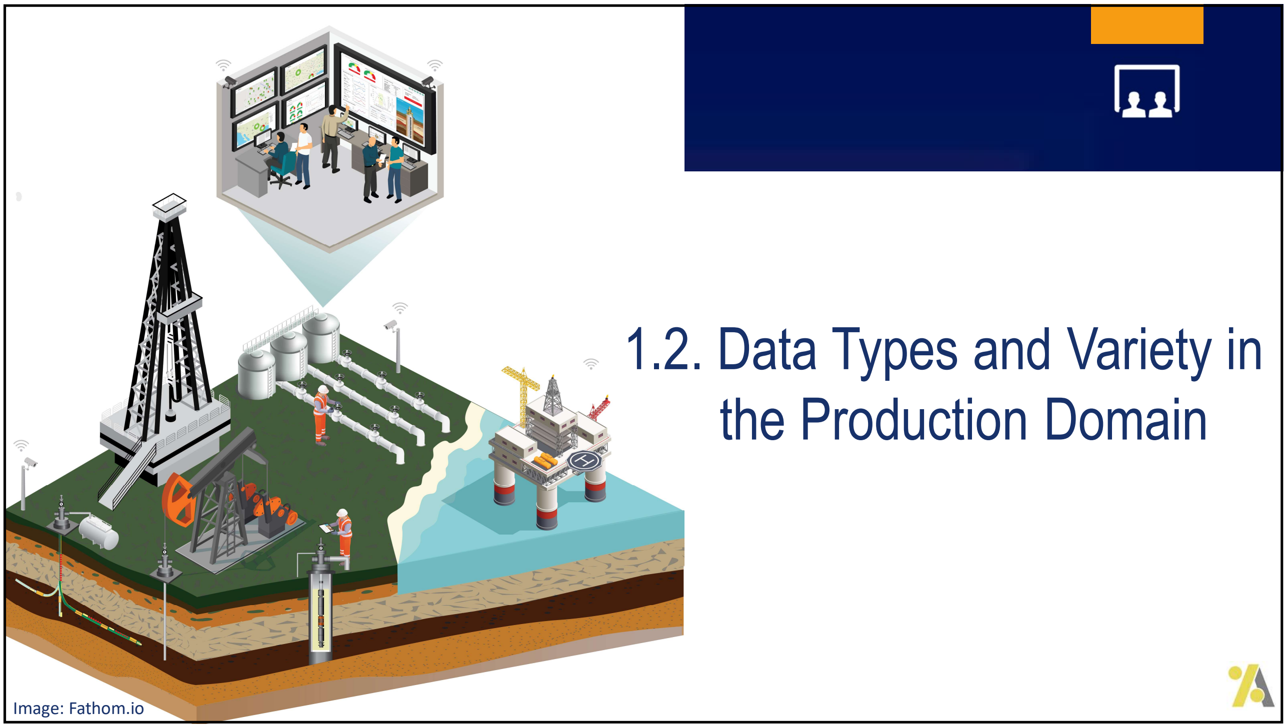
Combine these insights into a comprehensive conclusion that reveals the overall benefit to the organization.

25

Lessons Learned

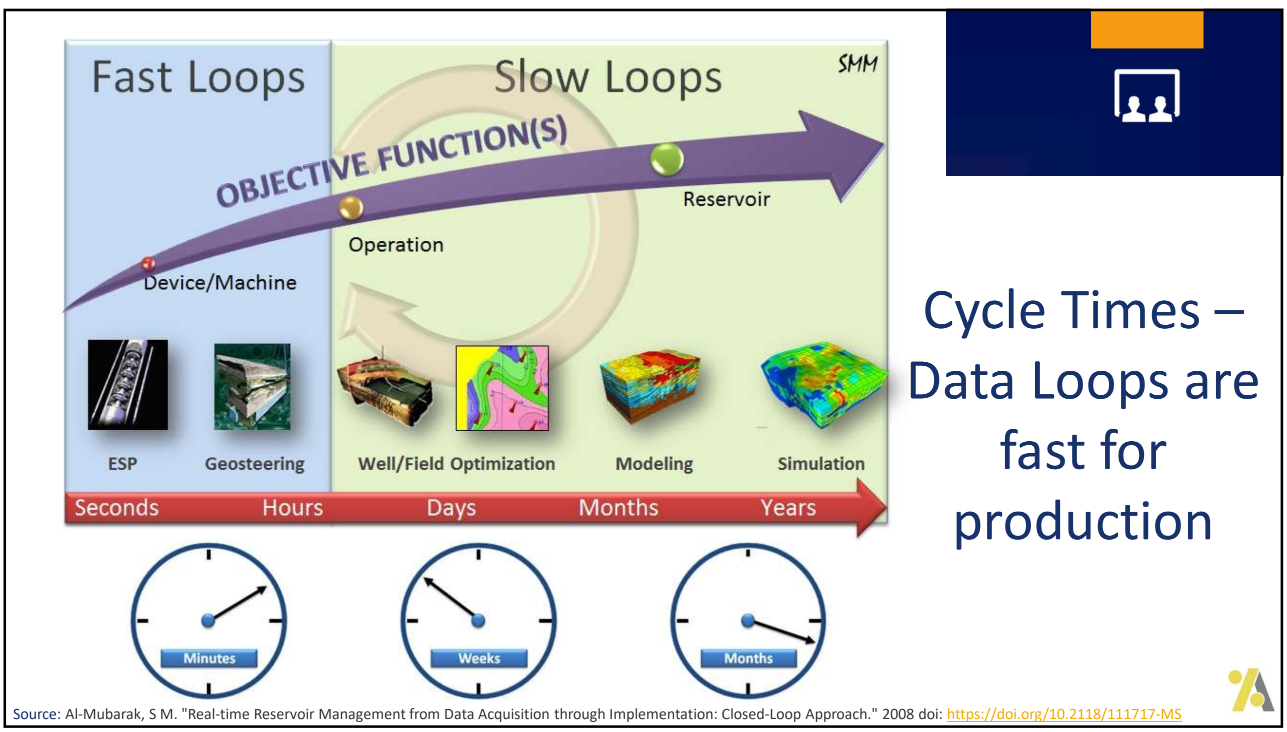
- **Investment**
  - In sensors is significant
  - In software is small
  - Software creates ROI
- **People**
  - Everyone must support and agree
  - Change management is important
- **Integration**
  - Effort is spent integrating IT systems
  - Departments must be integrated in expectations
- **Effort**
  - Mostly integration/organization
  - Some domain knowledge
  - Little data science
- **Conclusions**
  - There is much more to a data science project than data science!
  - It is not necessary for operators to have dedicated data science software developers – rather knowledgeable persons to organize, and manage the project and domain knowledge

26



1.2. Data Types and Variety in the Production Domain
Image: Fathom.io

27



Fast Loops
Slow Loops
SMM
OBJECTIVE FUNCTION(S)
Reservoir
Operation
Device/Machine
ESP
Geosteering
Well/Field Optimization
Modeling
Simulation
Seconds
Hours
Days
Months
Years
Minutes
Weeks
Months
Cycle Times – Data Loops are fast for production
Source: Al-Mubarak, S M. "Real-time Reservoir Management from Data Acquisition through Implementation: Closed-Loop Approach." 2008 doi: https://doi.org/10.2118/111717-MS

28

SENSE: Field Instrumentation – Surface Sensors, Meters

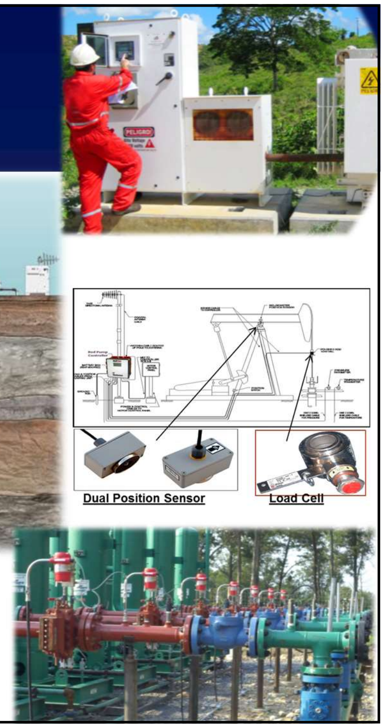
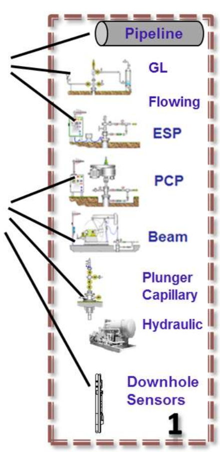


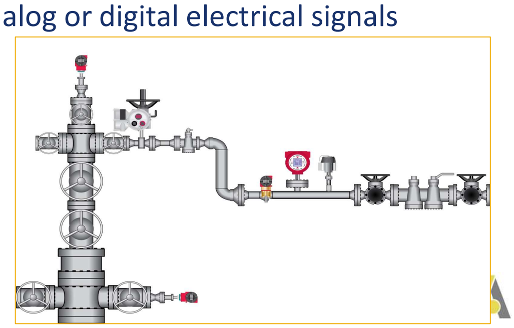
Image Source: Apergy, Borets, unknown

29

SENSE: Field Instrumentation – Surface Devices



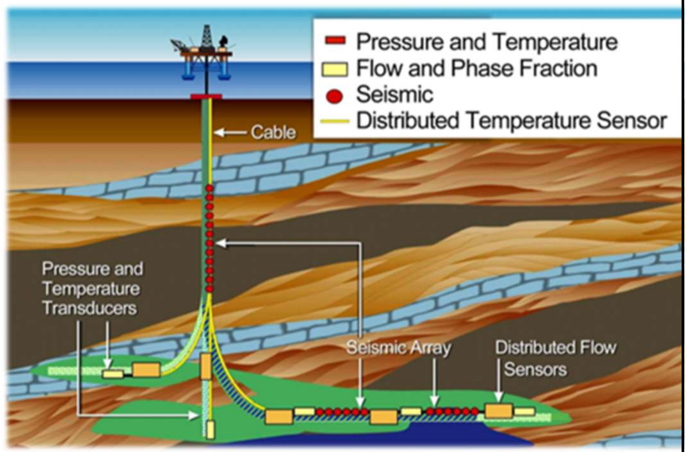
- Wells, flow lines and process measurement & control
  - **Pressure, temperature, P/T-differentials, load, position, torque, speed, vibration, corrosion/erosion rate, liquid level, water cut, gas/liquid flow rates, noise, visuals**, etc., to analog or digital electrical signals



30

SENSE: Field Instrumentation – Subsurface Sensors

- Permanent down-hole electronic or quartz gauges
  - Pressure, temperature, and vibration sensing using a Single conductor cable
- **Fiber Optics t**echnology
  - Distributed temperature sensing (DTS), readings every 0.5 meter
  - Distributed Acoustic Sensing
  - Pressure, temperature, flow, and seismic sensing using 1 to 4 fibers per cable
  - Higher quantity of measurement points; High volume of data generated
- Gauges can be multiplexed on the cable



31

31

SENSE & EXECUTE: Field Instrumentation – Wellsite Controllers, VFDs, Data Loggers

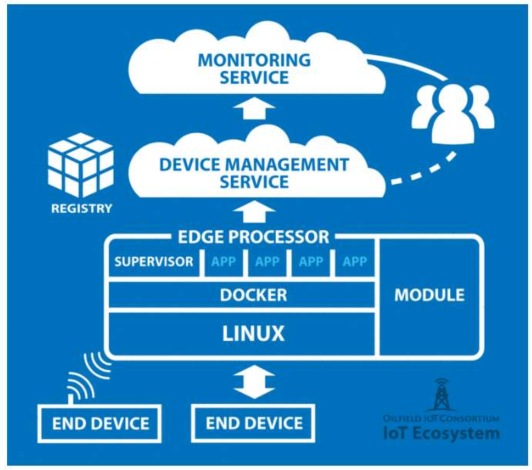
- Located close to well in a non-hazardous area with local display and operation.
- Connectivity to field hubs including data stores
  - Real-time data monitoring and transmittal to SCADA system
  - Remote management and adjustment of operating parameters
  - Slave-Master configurations (Internet of Things)
- Smart devices exhibit some autonomy based on measurements and built-in logic
  - Detect triggers and exercise control
    - Shut down if a pre-set limit violated
    - Auto-restart after a shutdown
    - Speed changes or Choke adjustment in response to changing well or surface conditions.

32

SENSE, STORE, EXECUTE, COMMUNICATE: Edge Devices

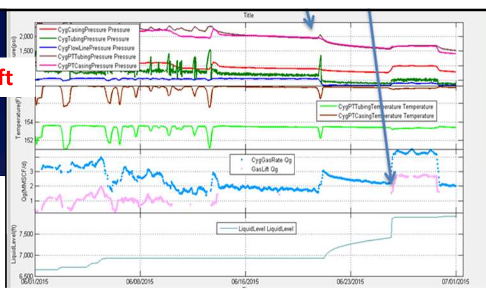
- Aggregate multiple devices to gain insight and perform more advanced control at site
- Open the door to high frequency data sampling for Machine Learning and Artificial Intelligence

Oilfield IoT Ecosystem

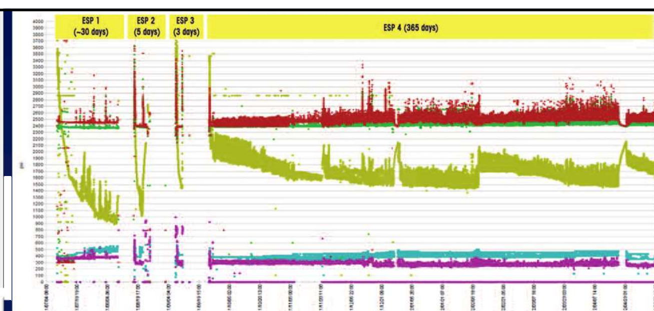


33

Gas Lift



Source: SPE 180949, Gonzalez, Chokshi, *et al.* Lessons Learned in Permian Gas Lift Shale Well: Dynamic Production Analysis with Downhole Gauge

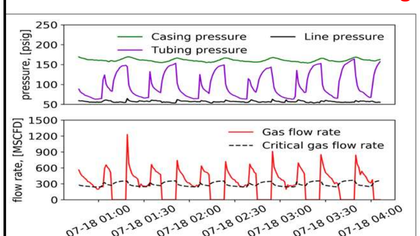


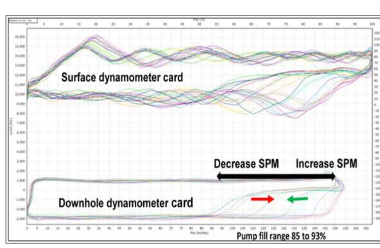
Source: MH Resources Corp, EPMagazine, April 2013

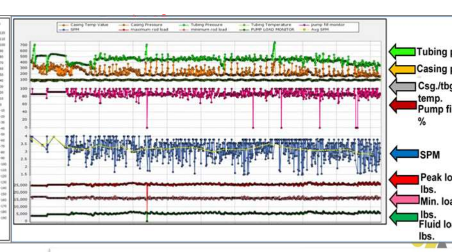
ESP

Types of Data – Time series Real-time – Point Measurements

Plunger Lift







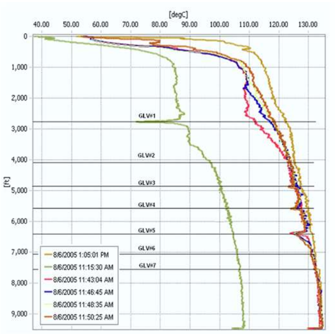
Source: Akhiiartdinov & Peyrera, Data Mining and Virtual Flow Metering in Plunger Lift Well, ALRDC GL Workshop, 2021

Source: SPE-181214, Clarke, F. and Malone, L.Sucker Rod Pumping in the Eagle Ford Shale Field Study.. 2016

34

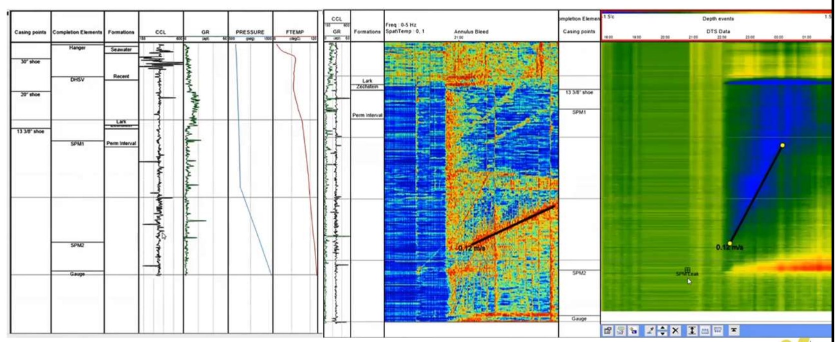
Types of Data: High Volume Time series Data – DTS, DAS

Slickline DTS to diagnose badly slugging Gas-lift well in Dubai



Source: Kumar SK, Dubai Petroleum, Gas Lift Diagnostics Using Distributed Temperature Sensor (DTS) System on Slick line, **ALRDC 2007 Gas-lift Workshop**

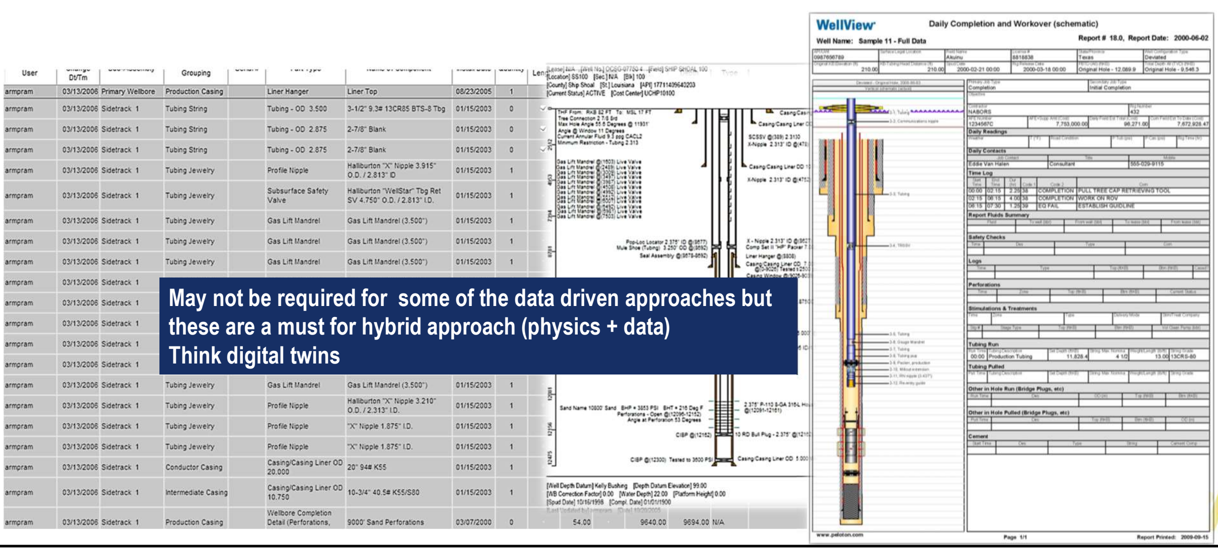
DTS + DAS for Packer / GLV Leak in North Sea



Source: Webster, M, Integration of DAS, DTS and conventional data, Sep 2020, https://www.youtube.com/watch?v=Rh4RwjhjFAA

35

Types of Data – Discrete events: Production, injection, well completion, reservoir, PVT, well activities, equipment changeovers, workovers, ……



36

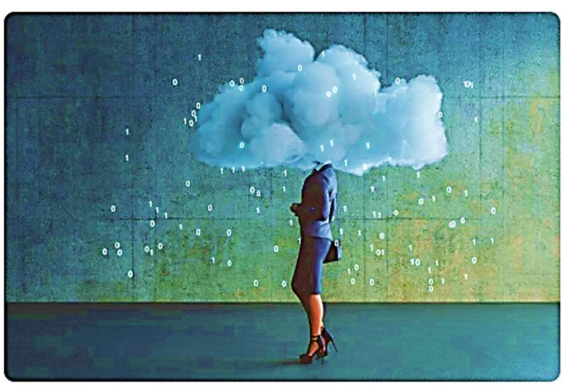
Data to Solve Artificial Lift Challenges

RRL	ESP	Gas-Lift	PCP	Plunger Lift	Production Data
Casing/Tubing/Line P, T PR Load, Displacement SPM Motor Amp, Voltage Stuffing Box level Fluid level Pumping Unit, Pump, Rods, Motor Specs	Casing/Tubing/Line P, T Pump Intake P, T Pump Discharge P, T Frequency/Speed Amp, Voltage Vibration I, V Imbalance I Leakage Fluid level Pump, Motor, VSD, Cable, Transformer, Power source Specs	Casing/Tubing/Line P, T Surface Gas Inj Rate Manifold Pressures Gradient Surveys Fluid Level GL valves' location specs, Inj Choke specs Valve performance	Casing/Tubing/Line P, T Speed/Frequency Pump Intake P, T Pump Discharge P, T Surface Flowrate Motor Amp, Voltage Motor Torque Fluid Level Pump, Motor, VSD, Power Source specs	Casing/Tubing/Line P, T Plunger arrival Gas Flow rate	Well Test, Choke Settings Well Config Reservoir PI/IPR Service Records Workover Records Teardown/DIFA Startup/shutdown data Facilities Status Injection (Corrosion, Surfactant)

Controller settings <u>Less Common:</u> Downhole gauge for GL, RRL, Plunger
<u>Rarely Available:</u> DTS, DAS, Multiphase Flow Meter Measurement

37

JPT [redacted]



DATA MANAGEMENT

All the Ways Bad Data Is Holding You Back

Whether inconsistent, incomplete, ambiguous, or just plain wrong, bad data is a big barrier to digital transformation.

May 31, 2023 • Engineering.com • Data Science and Digital Engineering

1.3. Data Processing Challenges

38

Implementation Caution : Field Realities

Data is not always what it seems: Need validation layer before analysis

Quality & Quantity Issues

- Key Measurements are missing; e.g., flow rate, pressure
- Measurements do not coordinate with each other; upstream << downstream pressure
- Measurements do not correspond to a physical model; Negative/Out of range Value, Flat-lined data, data clipping, outliers, discontinuities, repeated values.

Causes

- Older Field infrastructure
  - Analogue/old instrumentation conversion and need for new sensors
  - Communication: GSM/radio issues, cable/no cable, Bandwidth on GSM satellite. Line-of-sight
  - Legacy Data formats.. Non-Digital Archives
- Lack of information for some time intervals: loss of signal / data maintenance
- Malfunctioning measurement devices: Calibration
- Human errors in captures: Wrong records of equipment/ assets: wrong location / model

39

Incomplete Data & Bias in Models

- Incomplete data can lead to bias in AI
  - **Bias makes it challenging to develop AI that works for everyone.**
- No AI system is complex enough, nor dataset deep enough, to represent and understand humanity in all its diversity.

For example, if you were teaching AI to recognize shoes and only showed it imagery of sneakers, it wouldn't learn to recognize high heels, sandals or boots as shoes.

Source: "Making sense of artificial intelligence," https://atozofai.withgoogle.com/intl/en-GB/bias/

40

SPE ATCE SINCE 1924 | 16–18 October 2023 San Antonio, Texas, USA

1.3.1. Basic System for Cleaning, Filtration, Alarm & Regulation

- Flat-lined data
- Data clipping
- **Outliers**
- Discontinuities
- Repeated values.

Accutant Solutions

41

Data Cleaning Processes

Simple Criterion

- If data is above or below $(\mu \pm N \times \sigma)$, treat the data as outlier.
  - Works correctly only for simplest of the cases
  - May identify false outliers, may miss actual ones

Robust Algorithms

- Account for non-stationary processes
  - Using piecewise mean (μ) and STD (σ) over a small window helps deal with the non-stationary trend
  - Use estimates of mean and STD that are unbiased by potential outliers
    - Jackknife technique
  - Are points above or below (Mean ±N×STD) are true outliers?
    - Additional checks

Source: OTC 29642, McNeil S(2019), Real-time Cleaning of Time-series Data for a Floating System Digital Twin

42

42

Normal Data → Noise → Outliers

Outliers

Outlier Definitions:
- The normal data can be regarded as being generated by a model, and outliers can be regarded as deviations from this normal data.
- Outlier is an observation that **deviates so much from others** as to arouse suspicion that it was generated by a different mechanism.
- Outlier is an observation in a dataset which appears to be **inconsistent with the remainder** of that set of data.

How wrong can things go if we avoid detecting outliers?

- Outliers create a bias in the fit. Quality of fit decreases with increasing number of outliers

Case	EUR	% Difference
Analytical Model	7485.3	0
No outliers	7483.2	-0.03
10 outliers	8146.8	8.83
20 outliers	6612.7	-11.66
30 outliers	5593.3	-25.28

"Outliers can drastically reduce the diagnostic value and reliability of rate/pressure transient analysis workflows."

SPE-179958, Chaudhary & Lee (2016) Detecting and Removing Outliers in Production Data to Enhance Production Forecasting

43

Outlier Removal

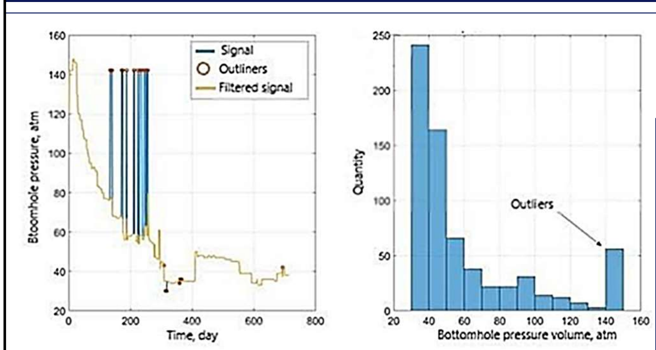


Figure 1—Removing outliers from bottomhole pressure measurement for production well A.

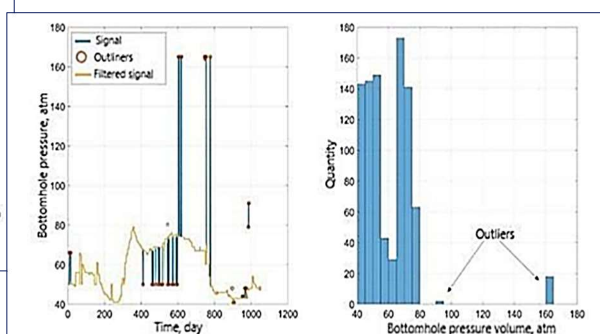
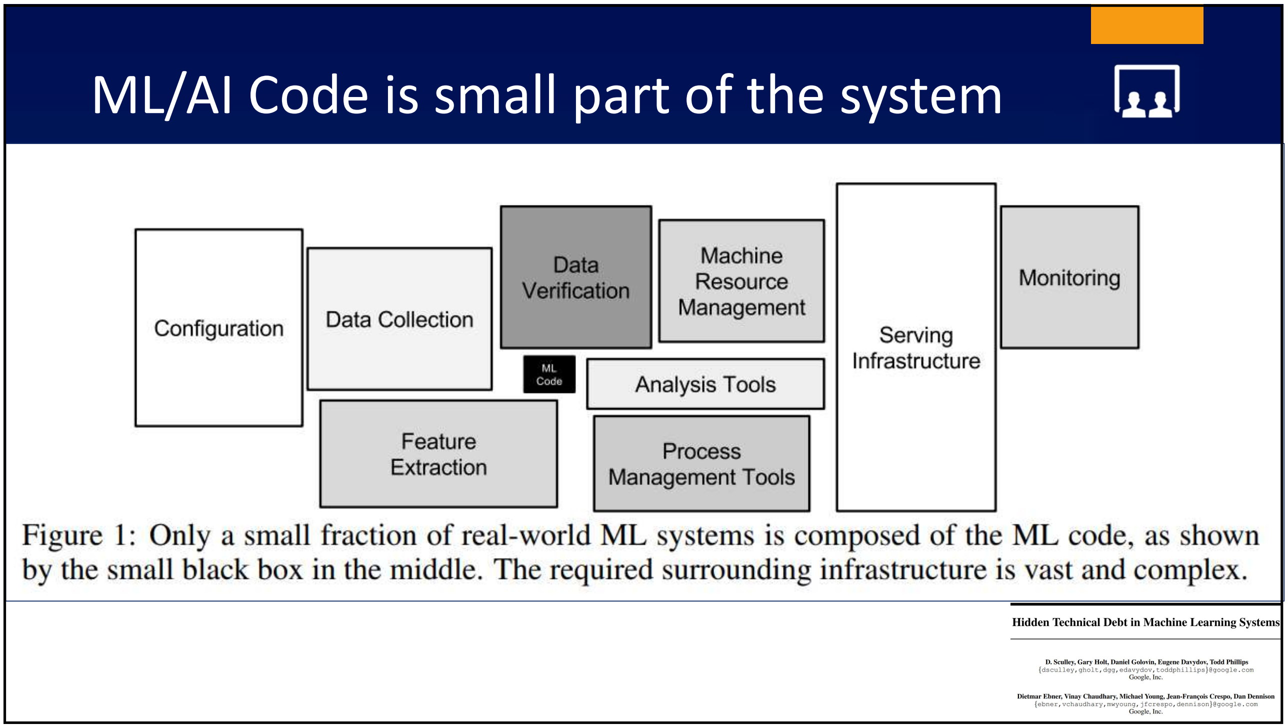


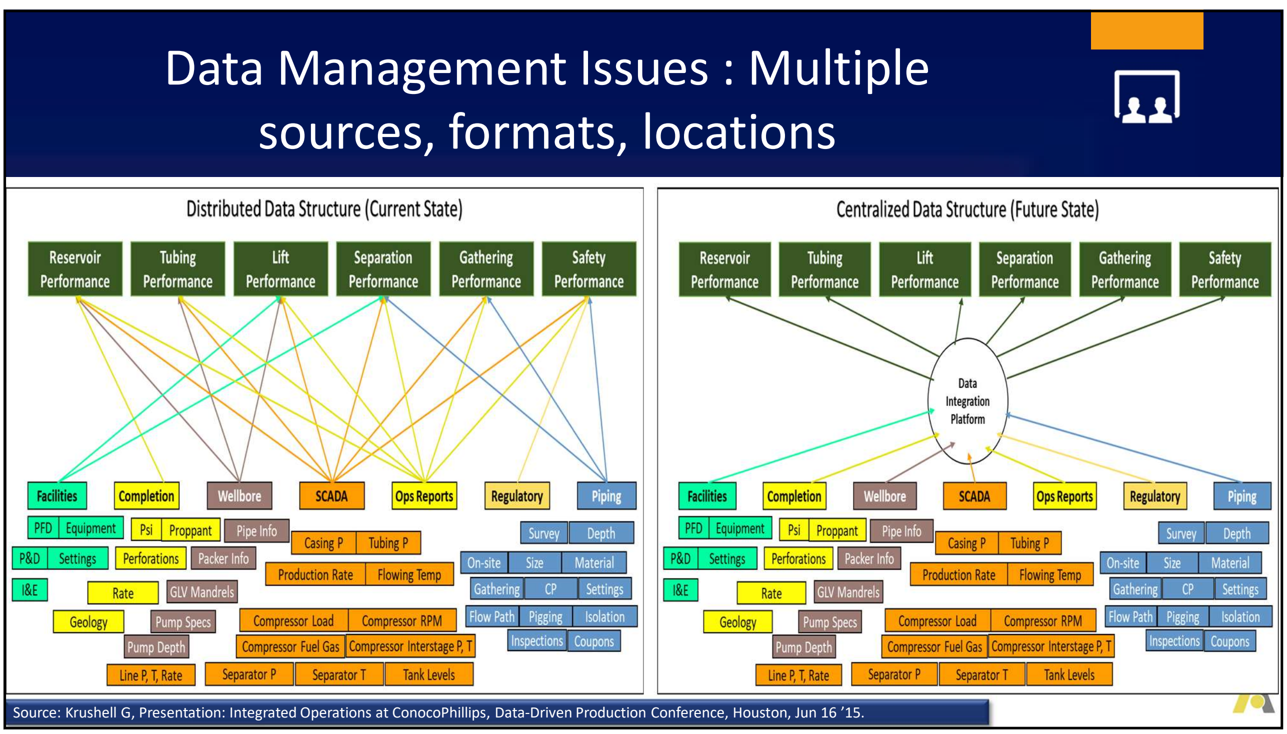
Figure 2—Removing outliers from bottomhole pressure measurement for production well B.

Source: Andrianova, A. et al. (2018). Application of Machine Learning for Oilfield Data Quality Improvement. doi:10.2118/191601-18RPTC-MS

44


ML/AI Code is small part of the system
Configuration
Data Collection
Data Verification
Machine Resource Management
Serving Infrastructure
Monitoring
ML Code
Analysis Tools
Feature Extraction
Process Management Tools
Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.
Hidden Technical Debt in Machine Learning Systems
D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
Google, Inc.
Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison
Google, Inc.

45


Data Management Issues : Multiple sources, formats, locations
Distributed Data Structure (Current State)
Reservoir Performance
Tubing Performance
Lift Performance
Separation Performance
Gathering Performance
Safety Performance
Facilities
Completion
Wellbore
SCADA
Ops Reports
Regulatory
Piping
PFD
Equipment
Psi
Proppant
Pipe Info
Casing P
Tubing P
Survey
Depth
P&D
Settings
Perforations
Packer Info
Production Rate
Flowing Temp
On-site
Size
Material
I&E
Rate
GLV Mandrels
Gathering
CP
Settings
Geology
Pump Specs
Compressor Load
Compressor RPM
Flow Path
Pigging
Isolation
Pump Depth
Compressor Fuel Gas
Compressor Interstage P, T
Inspections
Coupons
Line P, T, Rate
Separator P
Separator T
Tank Levels
Centralized Data Structure (Future State)
Data Integration Platform
Source: Krushell G, Presentation: Integrated Operations at ConocoPhillips, Data-Driven Production Conference, Houston, Jun 16 '15.

46

Unstructured Data: Ingestion Challenges

- For example, Field report(s) with
  - Well site pictures, videos showing condition of equipment/operations
  - Analog paper charts
  - Text messages
  - Email snippets
- Difficult to vectorize
- Metadata extraction
- Extra steps to perform arithmetic operations
- Difficult to plot: Plotting doesn't scale very well
- Systems to store files

47

Data Formats and Utilization

- Data formatting/representation Challenges in systems and software
  - a big challenge
    - real time data streams - units - / time zone issues
    - System interconnectivity hampered
- Energistics WITSML/PRODML/RESQML Data Standards -- https://www.energistics.org/
- OSDU technology-agnostic data platform -- https://osduforum.org/

48

1.4. Data Visualization

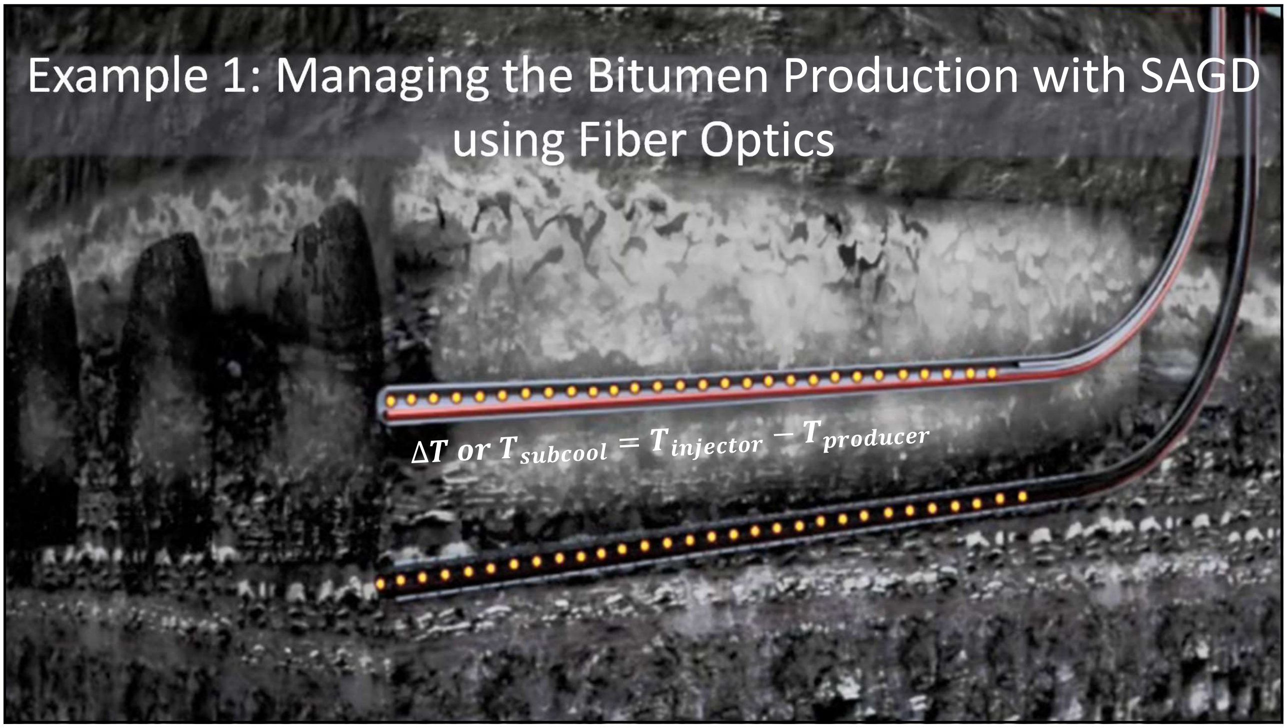
49

PowerBI visualization of Combined Data sets

Timeline visualization help

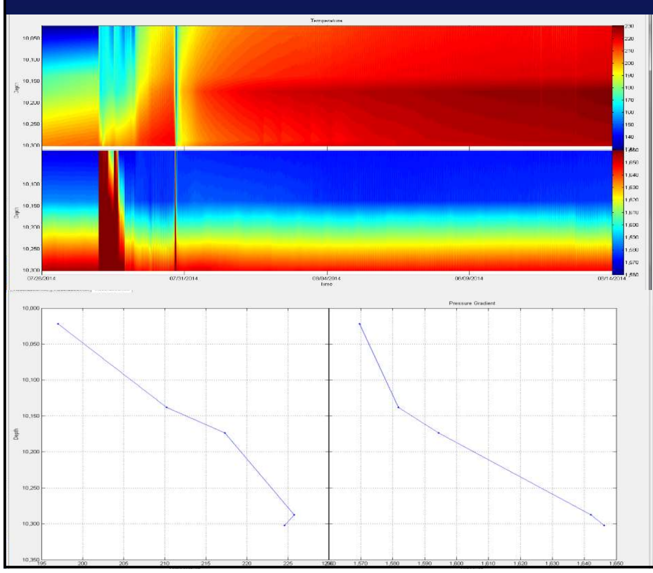
- Isolate an activity (or a group of activities) that contributed to uplift or deterioration in Well's health.
  - Did new completion (valve repositioned) improve production?
  - Increase in gas injection???
- Identify need for scheduling or deferring a remediation measure.
  - Is it time for another scale removal operation?
  - Did we conduct too many hot oiling jobs?
  - Should we move the well to a 'wet' header?

50



51

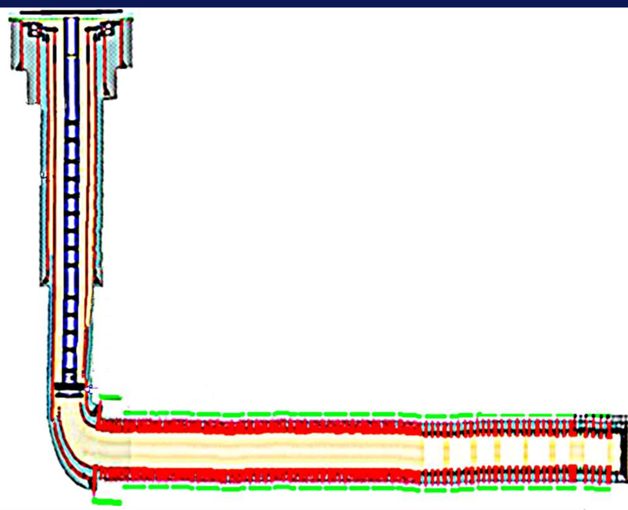
Example 2: Multiple point pressure/temperature in Gas-lift well

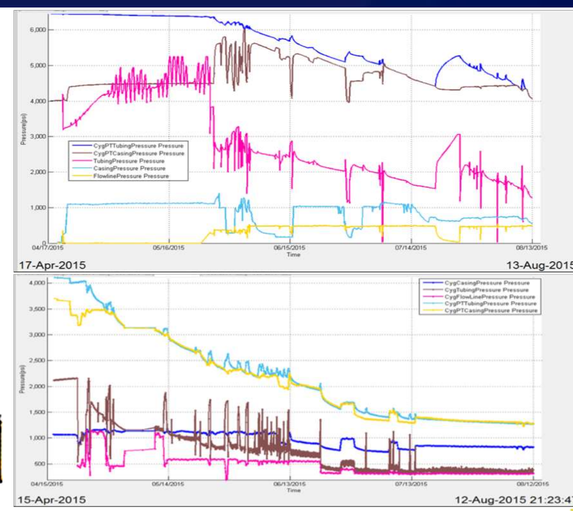


- Gulf Of Mexico vertical Gas Lift well
- Thermal EOR
- 5 P/T Gauges

52

Example 3 – Permian Condensate Well Gas-Lift Visualization



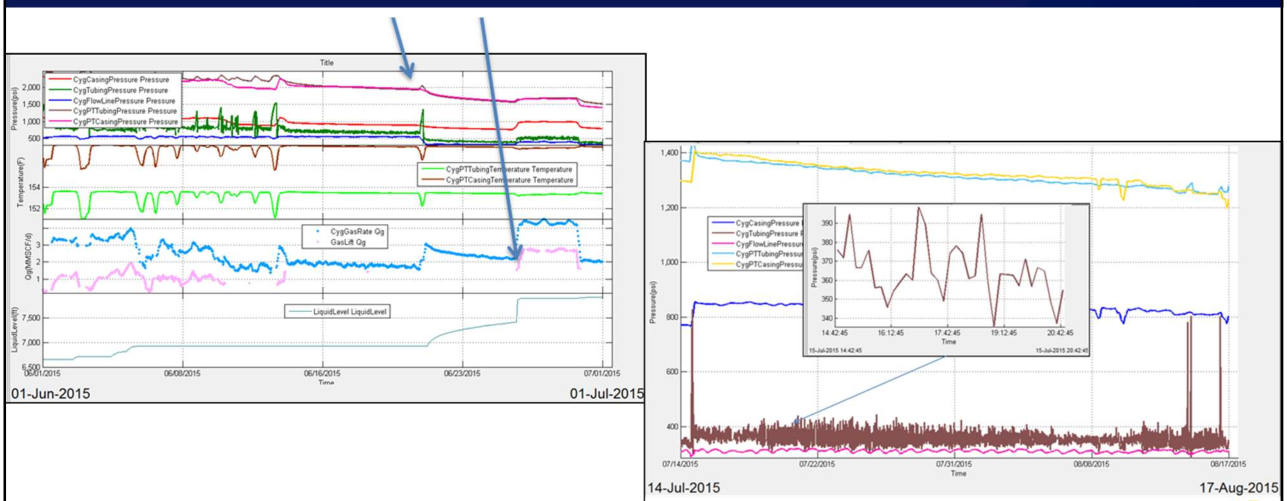


Source: SPE 180949, Gonzalez, Chokshi, *et al.* Lessons Learned in Permian Gas Lift Shale Well: Dynamic Production Analysis with Downhole Gauge, doi:10.2118/180949-MS

53

53

Example 3 (Contd): Permian Condensate Well : Visualization identifies gas-lift operational issues

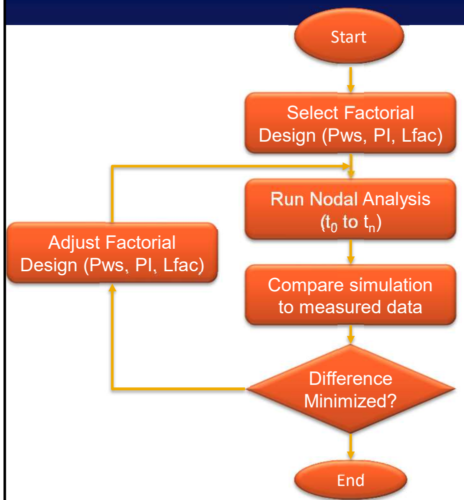


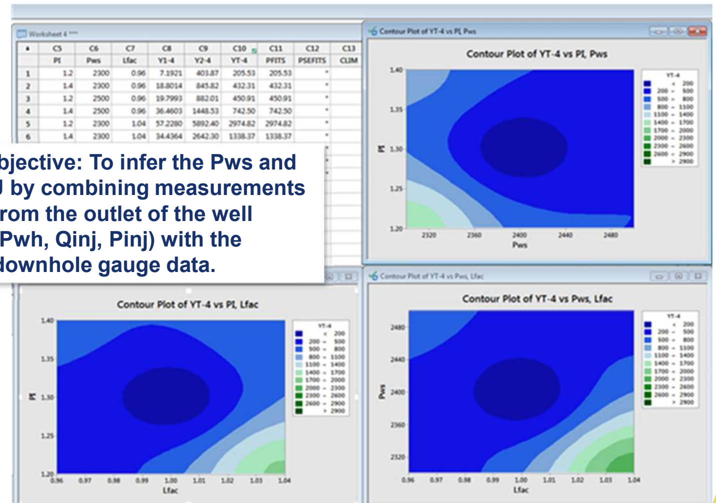
Source: SPE 180949, Gonzalez, Chokshi, *et al.* Lessons Learned in Permian Gas Lift Shale Well: Dynamic Production Analysis with Downhole Gauge, doi:10.2118/180949-MS

54

54

Example 3 (Contd): Permian Condensate Well : Real-time Measurements → Visualization → Predictive P_{ws}, J





Objective: To infer the Pws and J by combining measurements from the outlet of the well (Pwh, Qinj, Pinj) with the downhole gauge data.

Source: SPE 180949, Gonzalez, Chokshi, *et al.* Lessons Learned in Permian Gas Lift Shale Well: Dynamic Production Analysis with Downhole Gauge, doi:10.2118/180949-MS

55

55

56

Pan American ENERGY

2.0. A Brief & Incomplete Primer on ML/AI

Dr. Rajan Chokshi

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

Outline

- Data Science & Data Analytics
- AI and ML and Deep Learning
- Bias - Variance – Complexity Tradeoff
- Data Properties & Preparation
- Model Types
- Role of Domain Knowledge
- Training Model
- How good is my model?
- Toolsets
- Overview of a few AL/PO Case studies

2

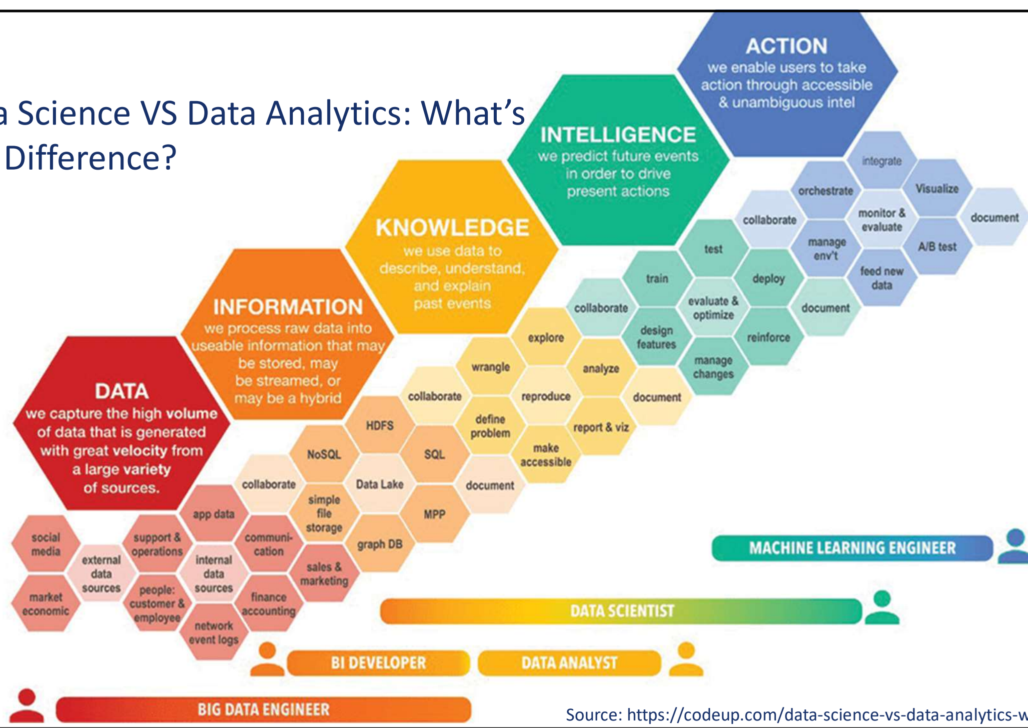
Data Science and Data Analytics: What's The Difference?

- Data Science
  - Multidisciplinary field for finding actionable insights from large sets of structured and unstructured data.
  - "Fixates on getting answers to the things we don't know, we don't know."
  - Incorporates computer science, predictive analytics, statistics, machine learning.
- Data Analytics
  - Focuses on processing and performing statistical analysis of existing datasets.
  - "Directed towards solving problems for questions we know we don't know the answers to."
  - Involves basic descriptive statistics, visualization and communication of conclusions.
- Data science has a wider scope compared to data analytics.
  - Data analytics is contained in data science and is one of the phases of the data science lifecycle.
  - What happens before and after analyzing the data is all part of data science.
    - Data cleansing, data preparation, analysis

Source: Liberty D., (2019) https://www.sisense.com/blog/data-science-vs-data-analytics/

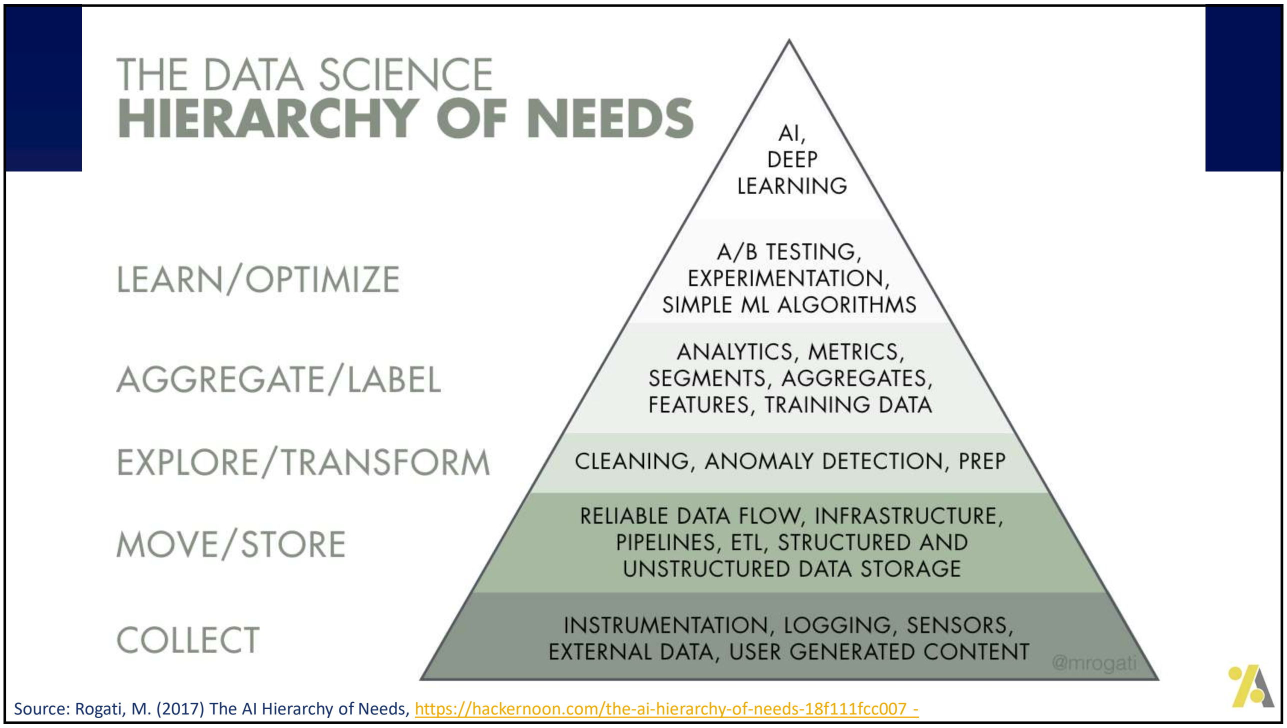
3

Data Science VS Data Analytics: What's The Difference?

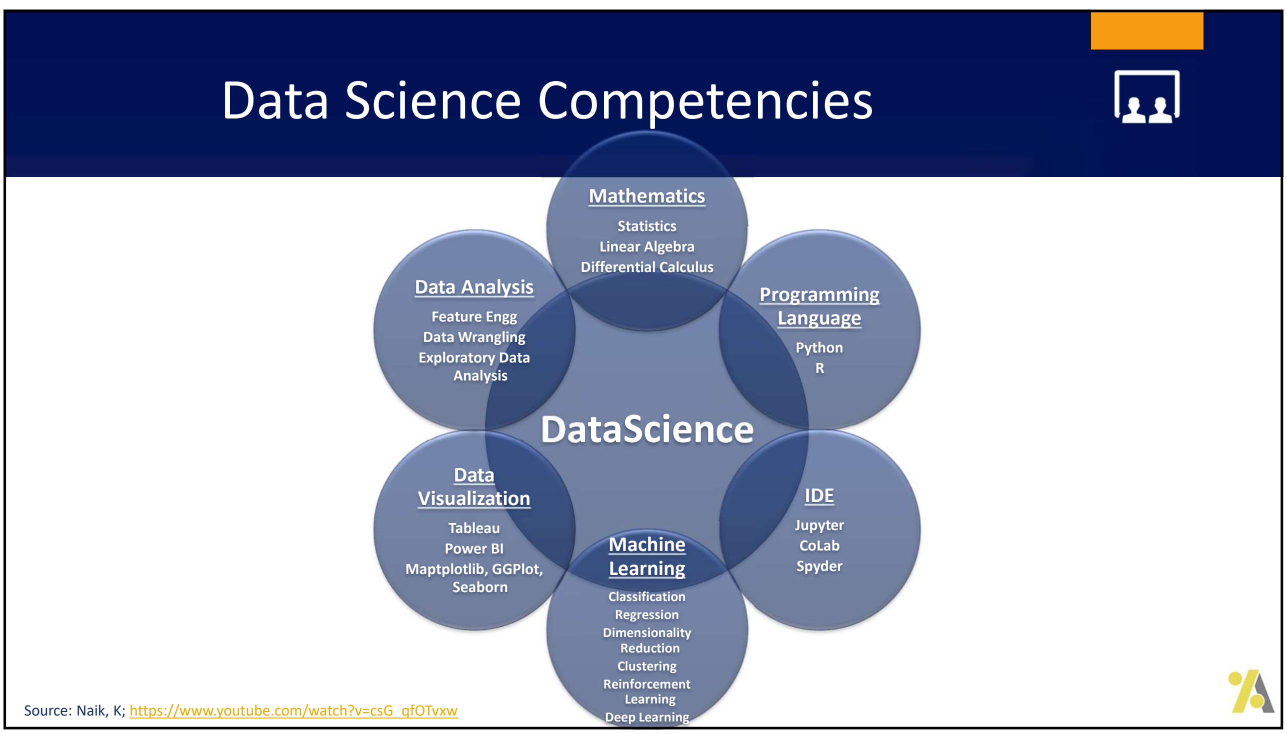


Source: https://codeup.com/data-science-vs-data-analytics-whats-the-difference/

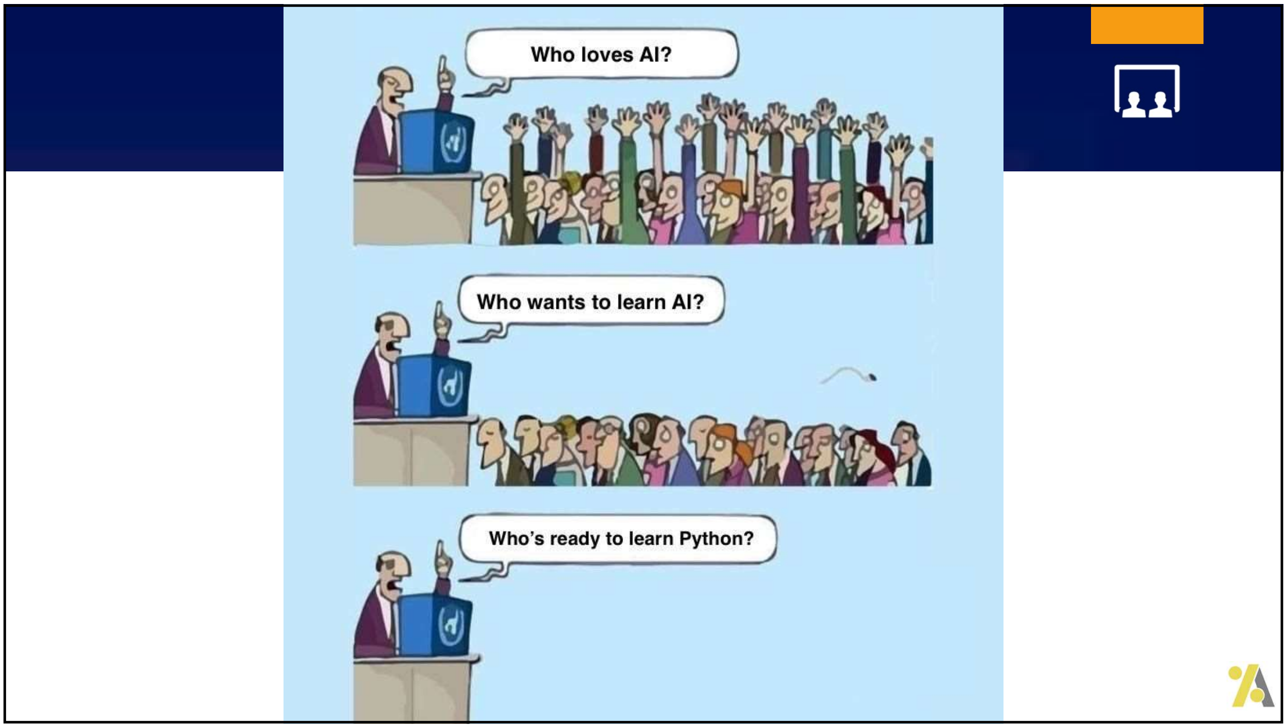
4


THE DATA SCIENCE
HIERARCHY OF NEEDS
AI,
DEEP
LEARNING
LEARN/OPTIMIZE
A/B TESTING,
EXPERIMENTATION,
SIMPLE ML ALGORITHMS
AGGREGATE/LABEL
ANALYTICS, METRICS,
SEGMENTS, AGGREGATES,
FEATURES, TRAINING DATA
EXPLORE/TRANSFORM
CLEANING, ANOMALY DETECTION, PREP
MOVE/STORE
RELIABLE DATA FLOW, INFRASTRUCTURE,
PIPELINES, ETL, STRUCTURED AND
UNSTRUCTURED DATA STORAGE
COLLECT
INSTRUMENTATION, LOGGING, SENSORS,
EXTERNAL DATA, USER GENERATED CONTENT
@mrogati
Source: Rogati, M. (2017) The AI Hierarchy of Needs, https://hackernoon.com/the-ai-hierarchy-of-needs-18f111fcc007 -

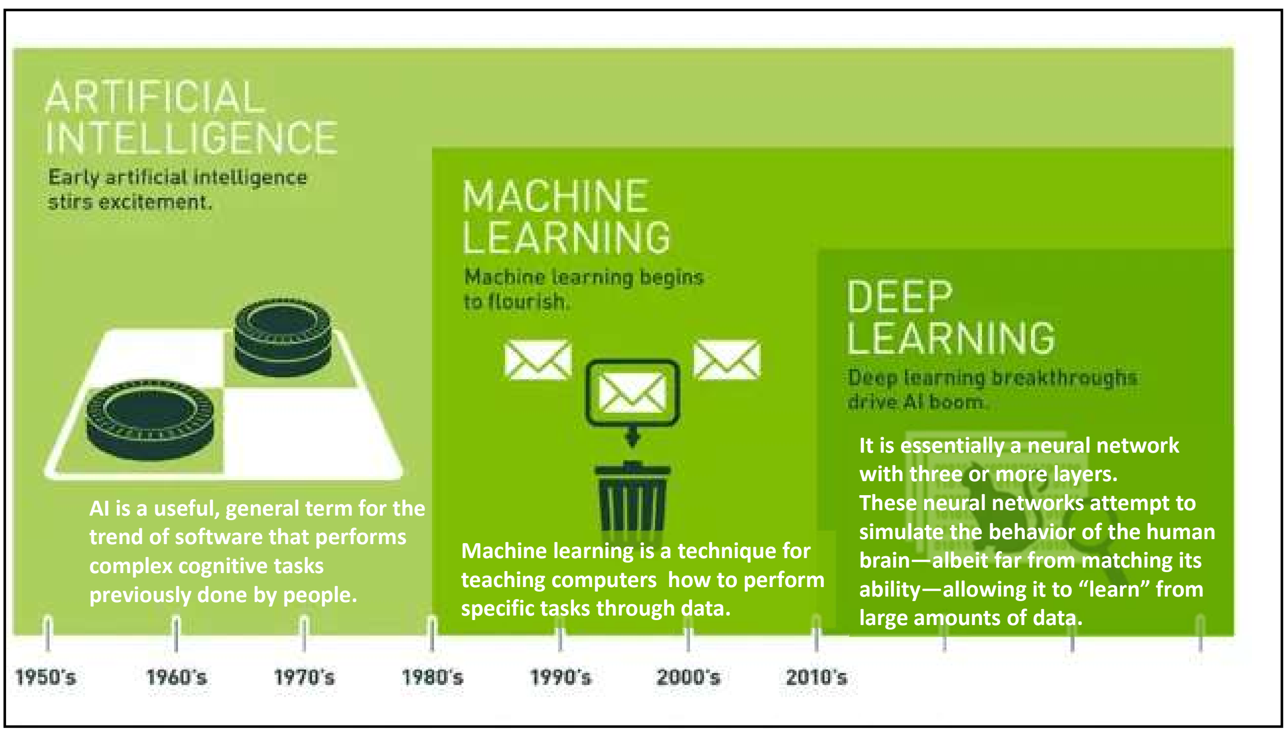
5


Data Science Competencies
Mathematics
Statistics
Linear Algebra
Differential Calculus
Data Analysis
Feature Engg
Data Wrangling
Exploratory Data
Analysis
Programming
Language
Python
R
DataScience
Data
Visualization
Tableau
Power BI
Maptplotlib, GGPlot,
Seaborn
IDE
Jupyter
CoLab
Spyder
Machine
Learning
Classification
Regression
Dimensionality
Reduction
Clustering
Reinforcement
Learning
Deep Learning
Source: Naik, K; https://www.youtube.com/watch?v=csG_qfOTvxw

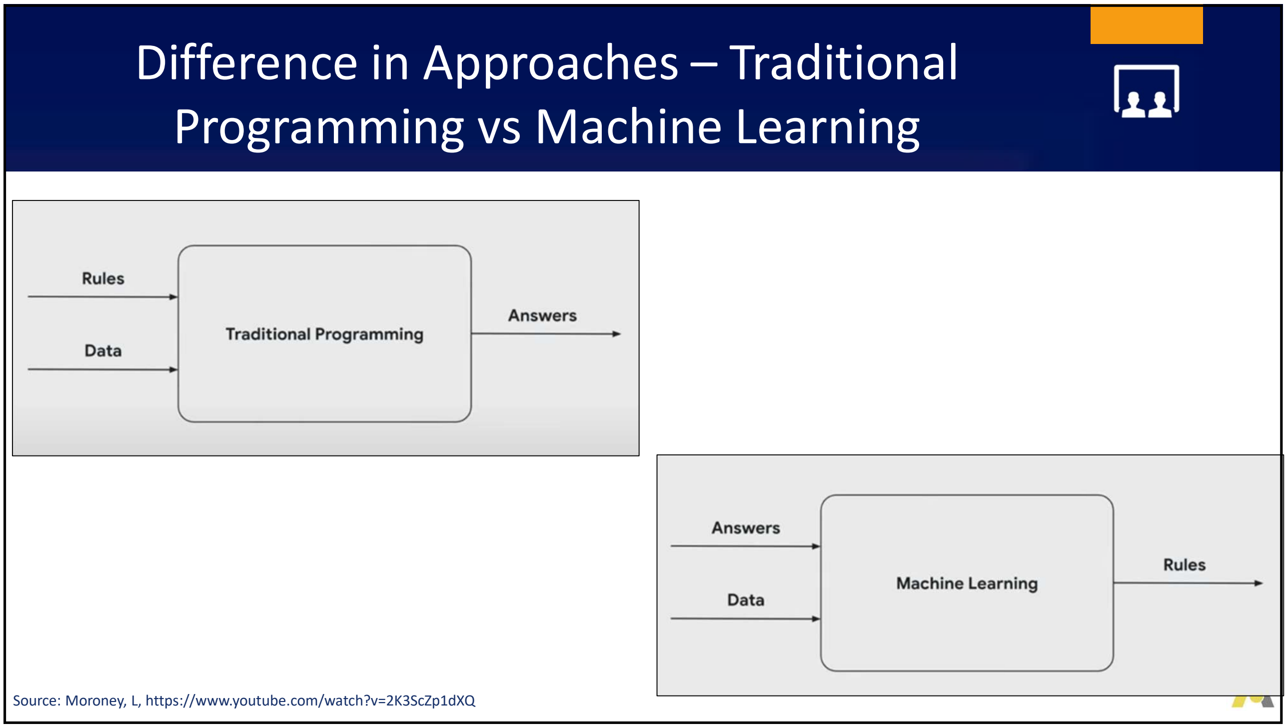
7


Who loves AI?
Who wants to learn AI?
Who's ready to learn Python?

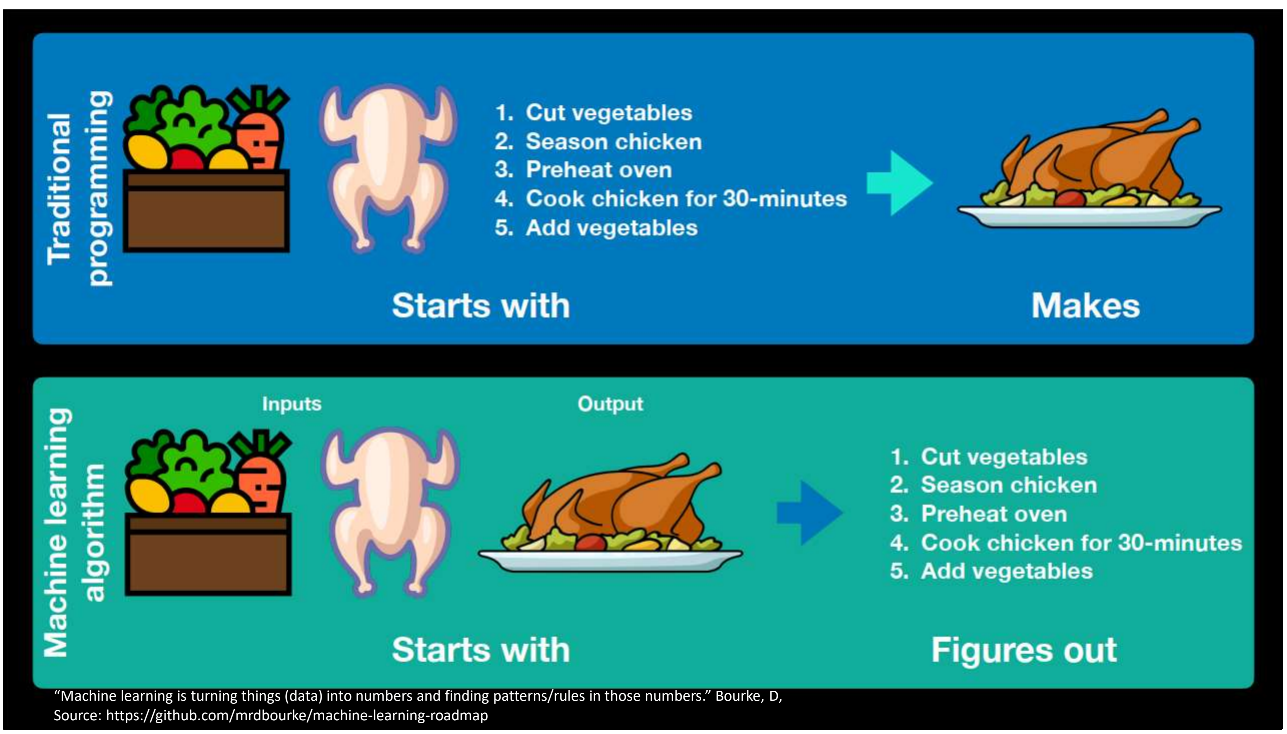
8


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
AI is a useful, general term for the trend of software that performs complex cognitive tasks previously done by people.
MACHINE LEARNING
Machine learning begins to flourish.
Machine learning is a technique for teaching computers how to perform specific tasks through data.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
It is essentially a neural network with three or more layers.
These neural networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to "learn" from large amounts of data.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

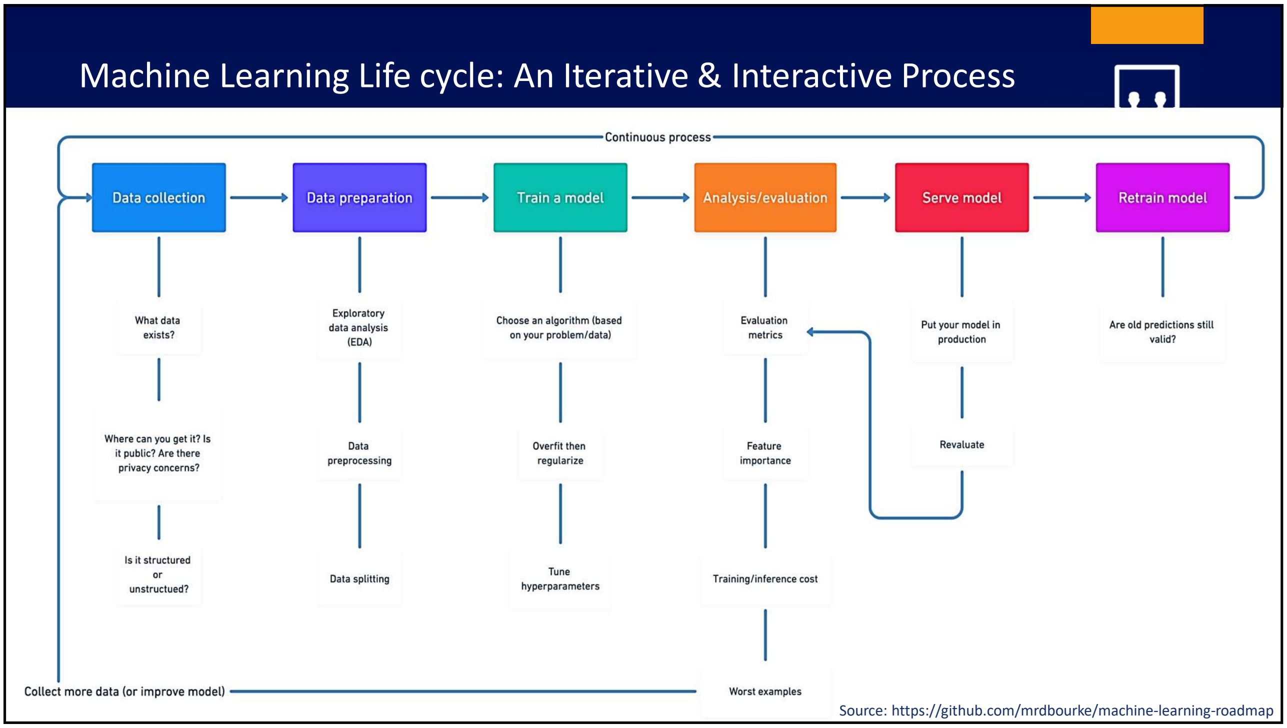
9


Difference in Approaches – Traditional Programming vs Machine Learning
Rules
Data
Traditional Programming
Answers
Answers
Data
Machine Learning
Rules
Source: Moroney, L, https://www.youtube.com/watch?v=2K3ScZp1dXQ

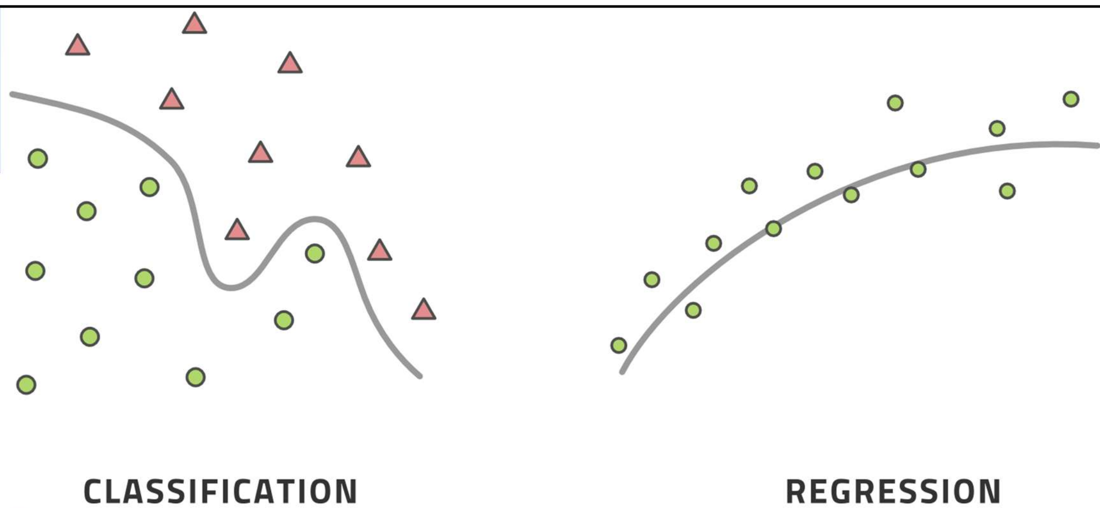
10


Traditional programming
1. Cut vegetables
2. Season chicken
3. Preheat oven
4. Cook chicken for 30-minutes
5. Add vegetables
Starts with
Makes
Machine learning algorithm
Inputs
Output
1. Cut vegetables
2. Season chicken
3. Preheat oven
4. Cook chicken for 30-minutes
5. Add vegetables
Starts with
Figures out
"Machine learning is turning things (data) into numbers and finding patterns/rules in those numbers." Bourke, D,
Source: https://github.com/mrdbourke/machine-learning-roadmap

11

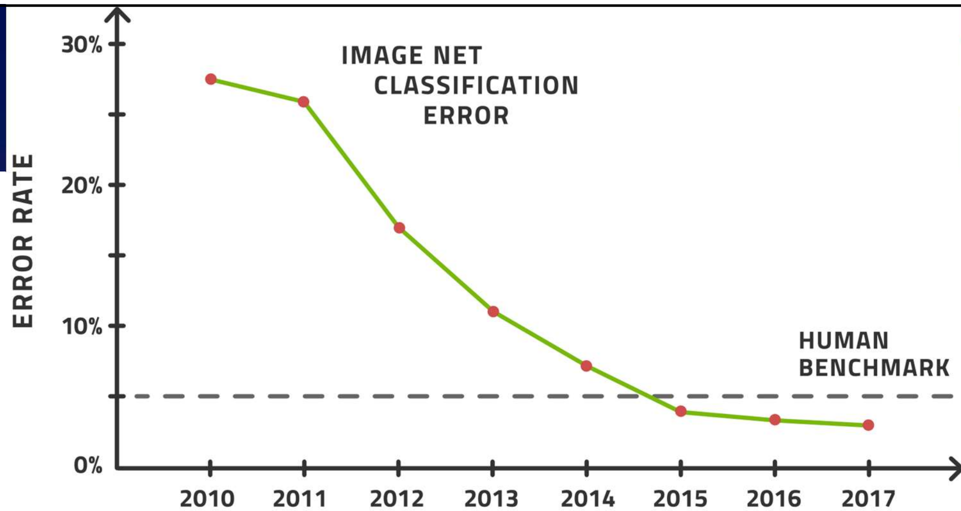


12



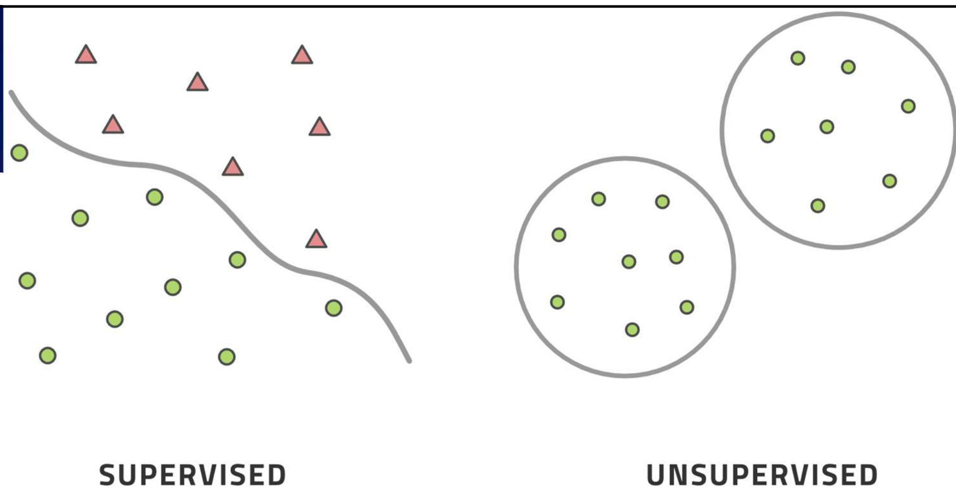
Machine learning tasks are either classification (which group does a data point belong to?) or regression (what numerical value does this variable have?).

13



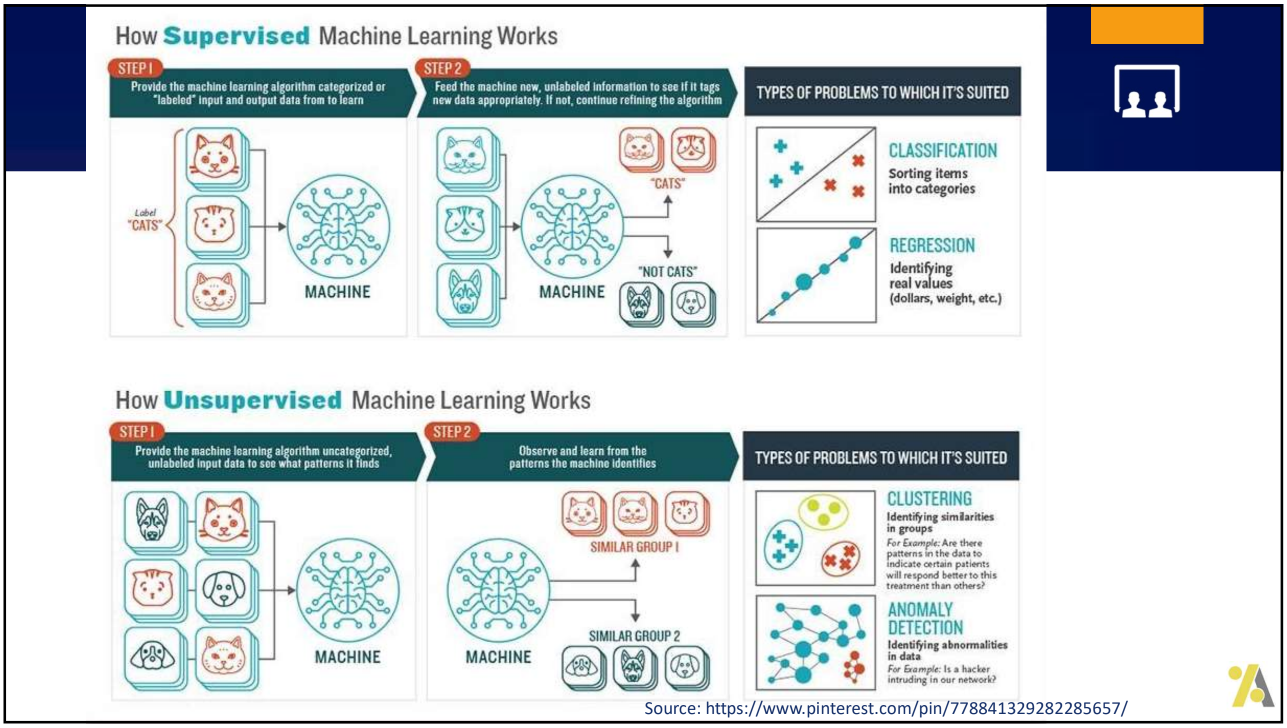
The accuracy of machine learning models began to rival human performance in 2014-2015. The drivers for accuracy improvement are: More data, more processing power, better algorithms.

14

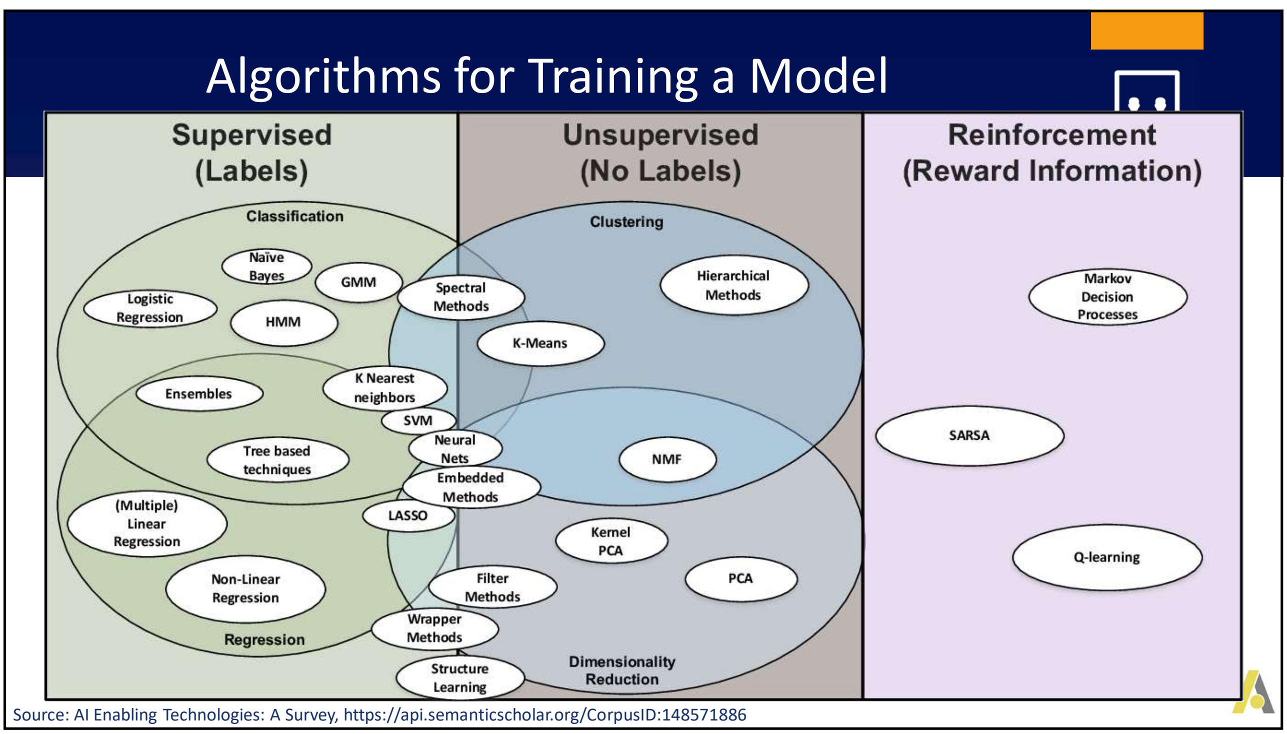


If the desired outcome of the computation is available to make the model, it is a supervised method. Otherwise, it is unsupervised. Different models and algorithms are used in these scenarios.

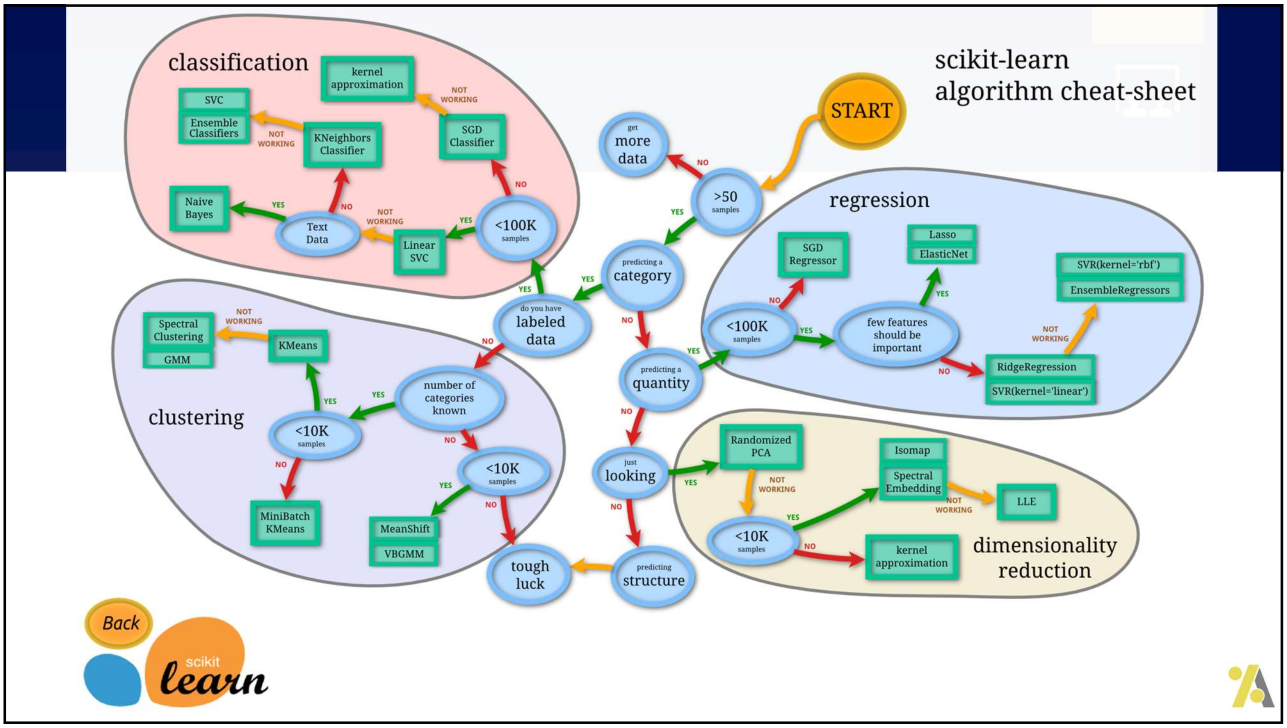
15


How Supervised Machine Learning Works
STEP 1
Provide the machine learning algorithm categorized or "labeled" input and output data from to learn
Label "CATS"
MACHINE
STEP 2
Feed the machine new, unlabeled information to see if it tags new data appropriately. If not, continue refining the algorithm
MACHINE
"CATS"
"NOT CATS"
TYPES OF PROBLEMS TO WHICH IT'S SUITED
CLASSIFICATION
Sorting items into categories
REGRESSION
Identifying real values (dollars, weight, etc.)
How Unsupervised Machine Learning Works
STEP 1
Provide the machine learning algorithm uncategorized, unlabeled input data to see what patterns it finds
MACHINE
STEP 2
Observe and learn from the patterns the machine identifies
MACHINE
SIMILAR GROUP 1
SIMILAR GROUP 2
TYPES OF PROBLEMS TO WHICH IT'S SUITED
CLUSTERING
Identifying similarities in groups
For Example: Are there patterns in the data to indicate certain patients will respond better to this treatment than others?
ANOMALY DETECTION
Identifying abnormalities in data
For Example: Is a hacker intruding in our network?
Source: https://www.pinterest.com/pin/778841329282285657/

16

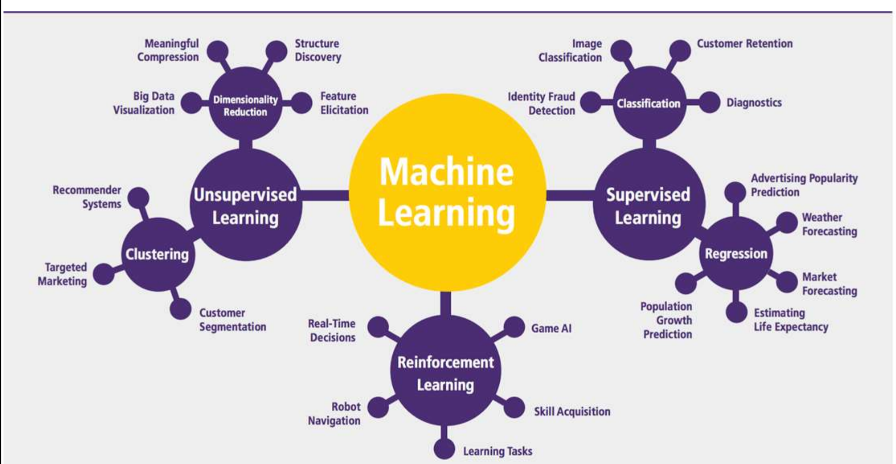

Algorithms for Training a Model
Supervised (Labels)
Unsupervised (No Labels)
Reinforcement (Reward Information)
Classification
Naïve Bayes
GMM
Logistic Regression
HMM
Spectral Methods
Clustering
Hierarchical Methods
K-Means
Ensembles
K Nearest neighbors
SVM
Neural Nets
Tree based techniques
Embedded Methods
NMF
(Multiple) Linear Regression
LASSO
Kernel PCA
PCA
Non-Linear Regression
Filter Methods
Wrapper Methods
Regression
Structure Learning
Dimensionality Reduction
Markov Decision Processes
SARSA
Q-learning
Source: AI Enabling Technologies: A Survey, https://api.semanticscholar.org/CorpusID:148571886

17



18

Selection of ML Algorithm Depends on Problem Type

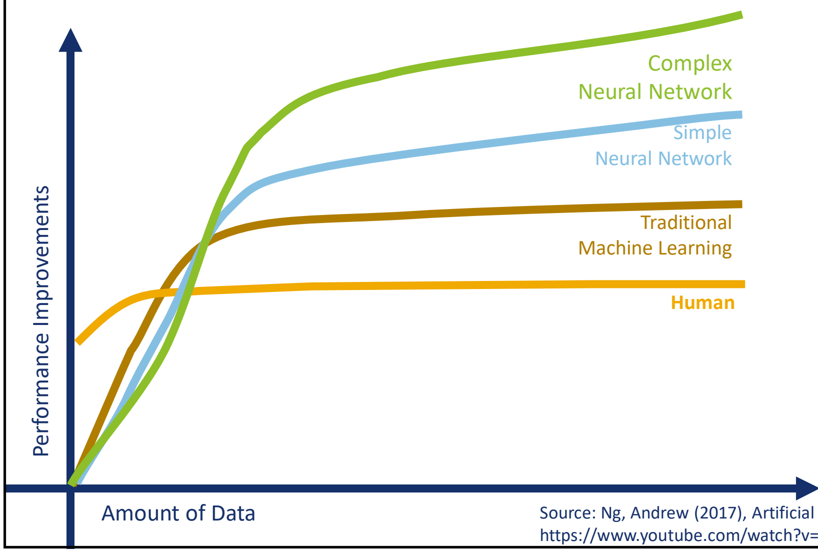


Question:
- List problems in artificial lift and production analysis
- Decide which category of ML-algorithms might work.

Source: Bag S "Which Machine learning Algorithm Should You Use By Problem Type?, "https://bit.ly/3feDBwO

19

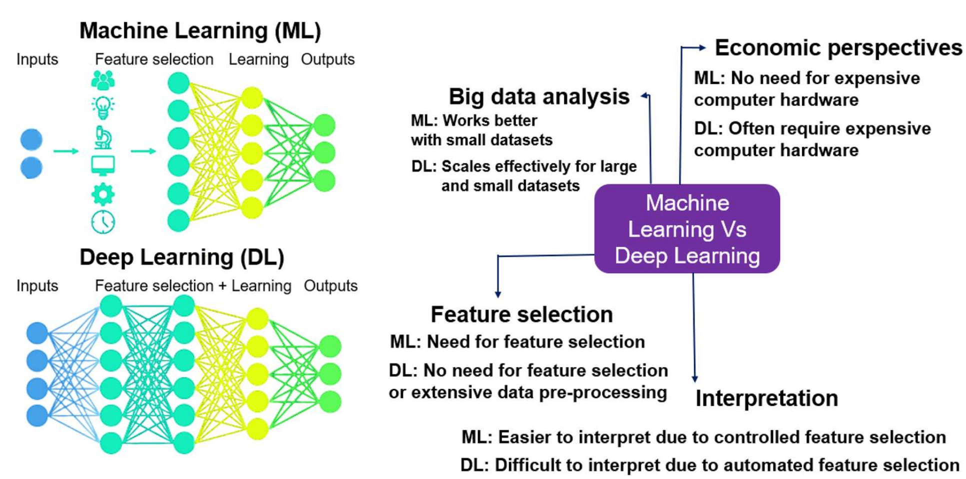
Deep Learning Improves as Data Grows



Source: Ng, Andrew (2017), Artificial Intelligence is the New Electricity, https://www.youtube.com/watch?v=21EiKfQYZXc

20

Deep Learning versus Machine Learning

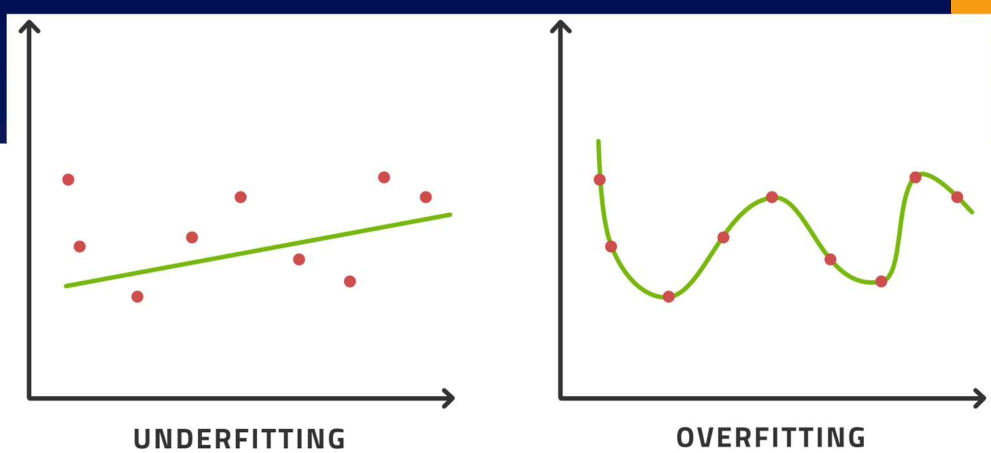


Source: Barjouei et al. Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes. J Petrol Explor Prod Technol 11, 1233–1261 (2021). https://doi.org/10.1007/s13202-021-01087-4

21

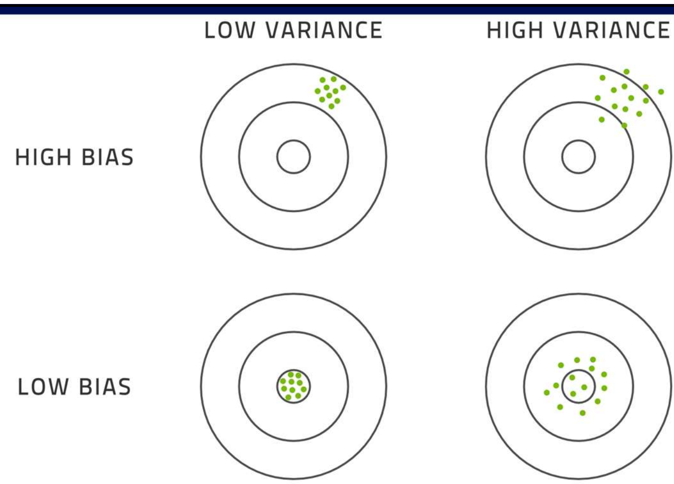
Bias-Variance-Complexity Trade-Off

22



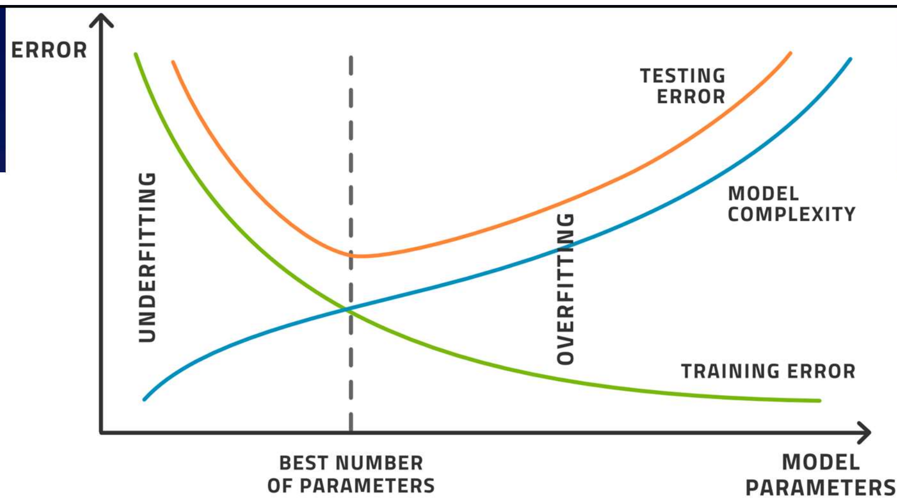
When the model is too simple (left) it cannot represent the data well. When the model is too complex (right) it can memorize the data and cannot generalize

23



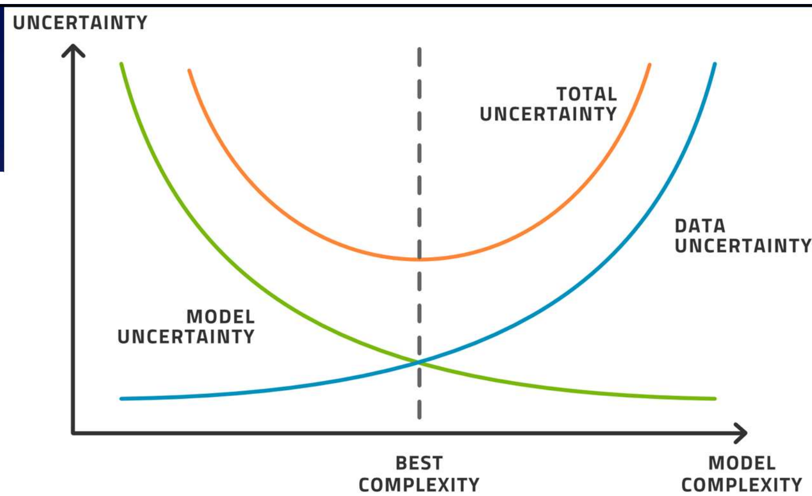
Changing the equation and the number of its parameters can change the bias and variance of the model's performance. An optimal balance must be found.

24



As the model parameters increase, the training error goes down (overfitting) and the test error eventually goes up (bias). The optimal model size must be capable of representing the situation.

25



Models with more parameters need more data in order to adjust the parameters well.

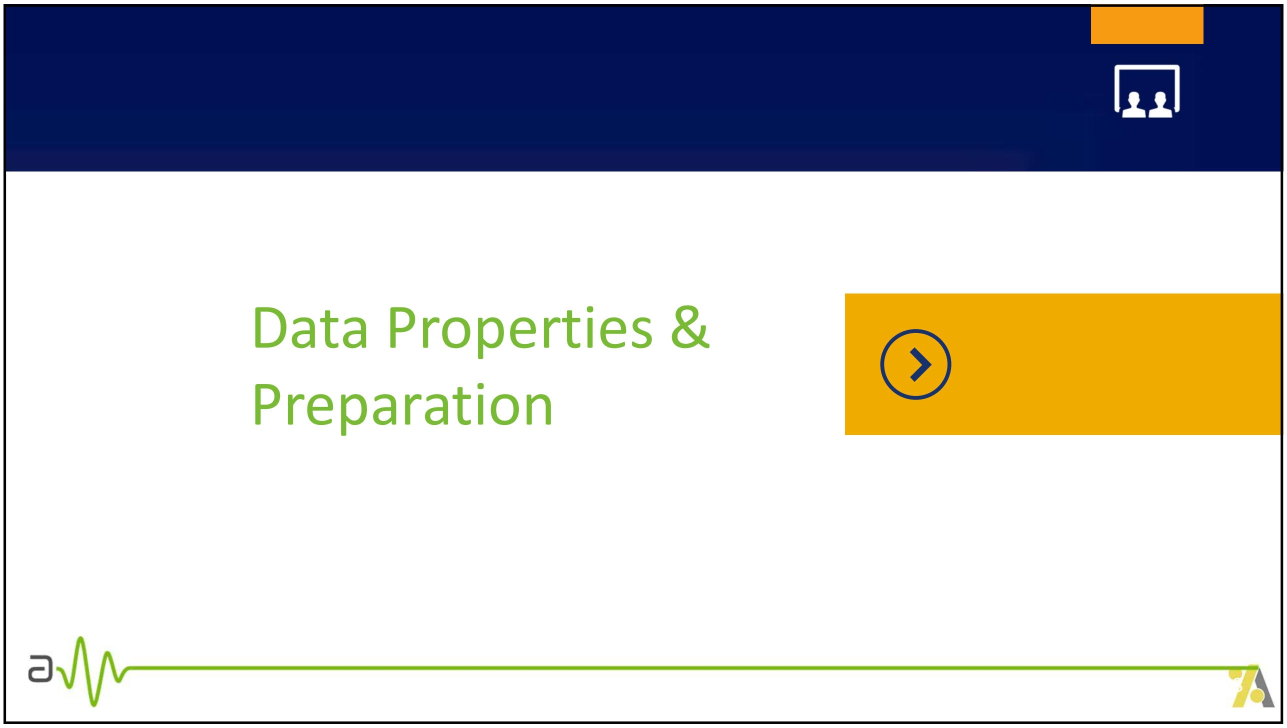
26

Model too small
It cannot represent the situation
=> **Variance**

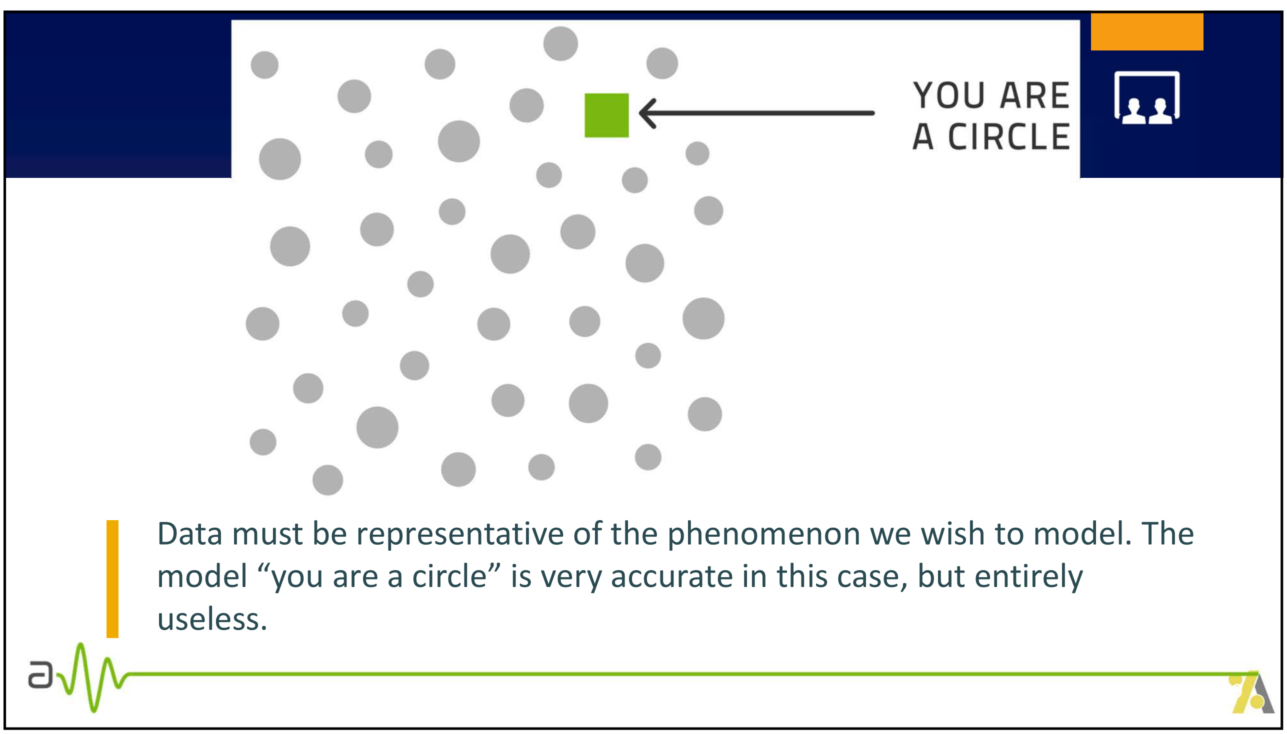
Model too big
It will overfit and not generalize =>
Bias

Often the solution is to get more data. Most of the advances of deep learning are due to vastly increased data sizes. If this cannot be done, we must sacrifice variance!

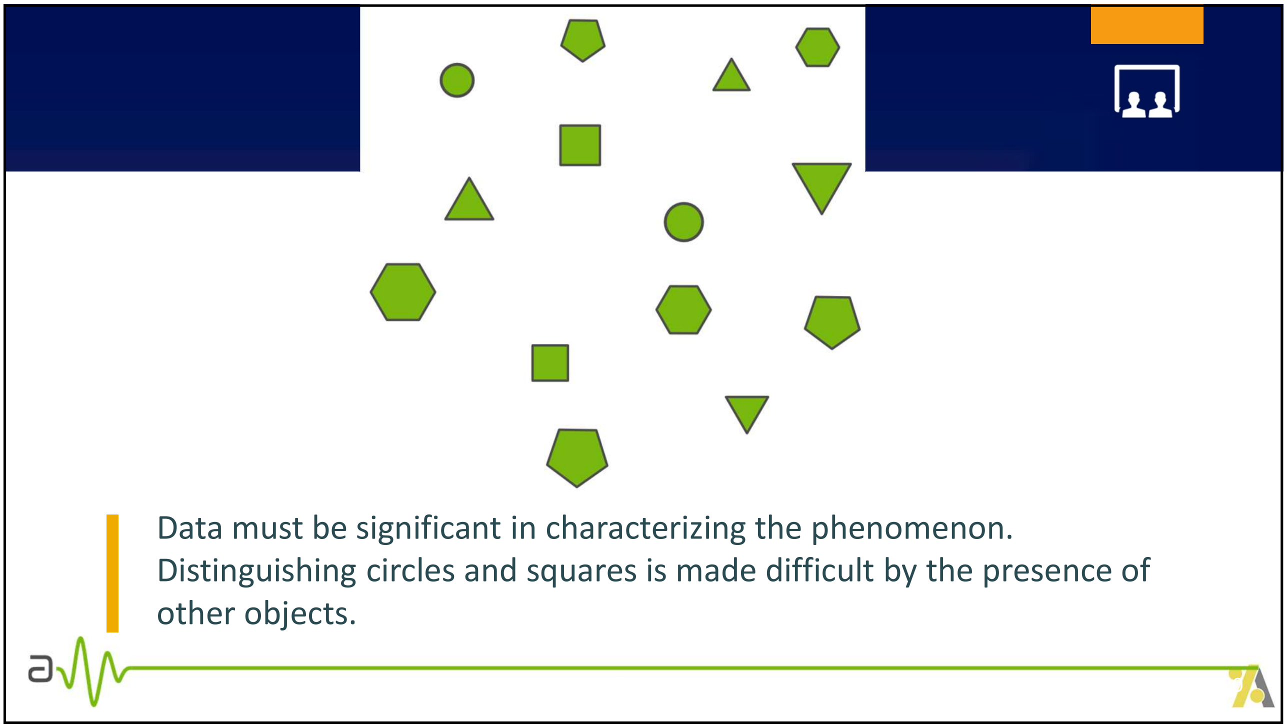
27


Data Properties &
Preparation

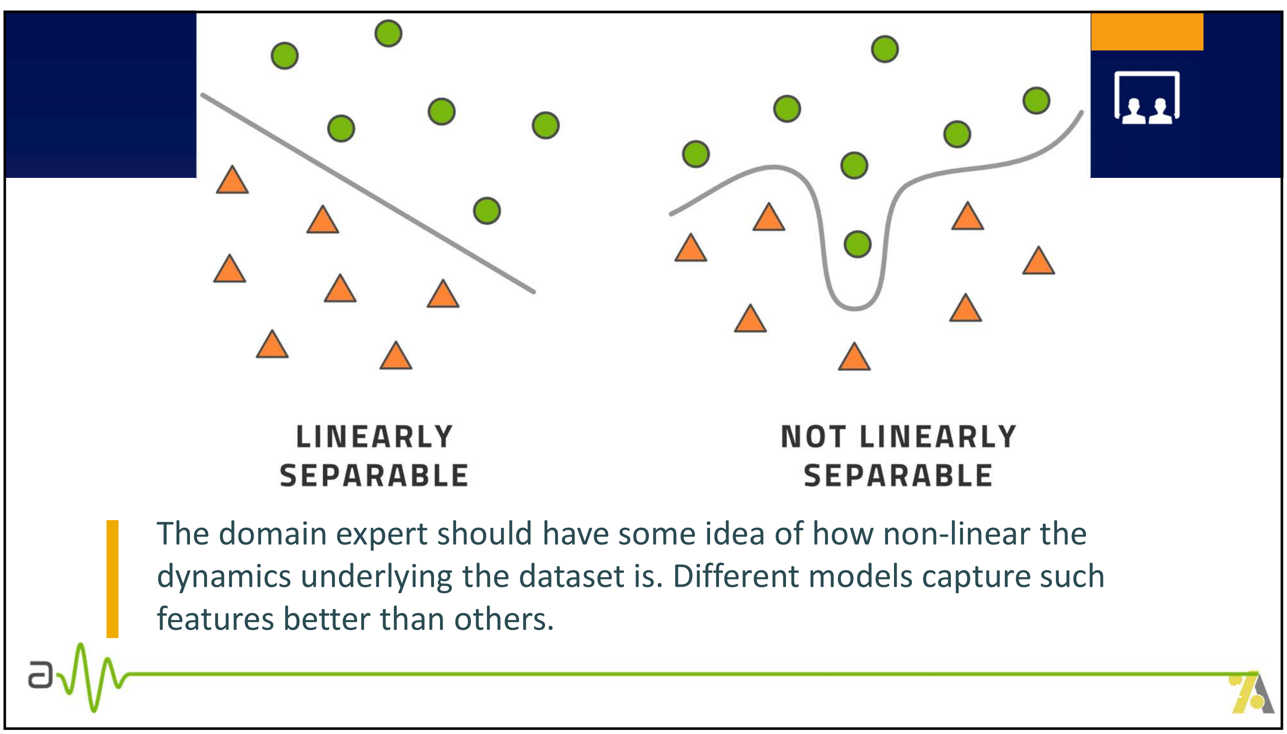
28


YOU ARE
A CIRCLE
Data must be representative of the phenomenon we wish to model. The model "you are a circle" is very accurate in this case, but entirely useless.

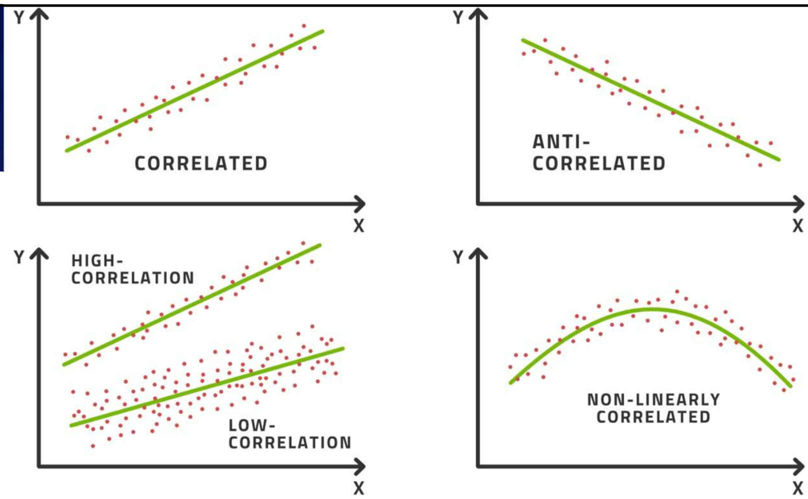
29


Data must be significant in characterizing the phenomenon. Distinguishing circles and squares is made difficult by the presence of other objects.

30

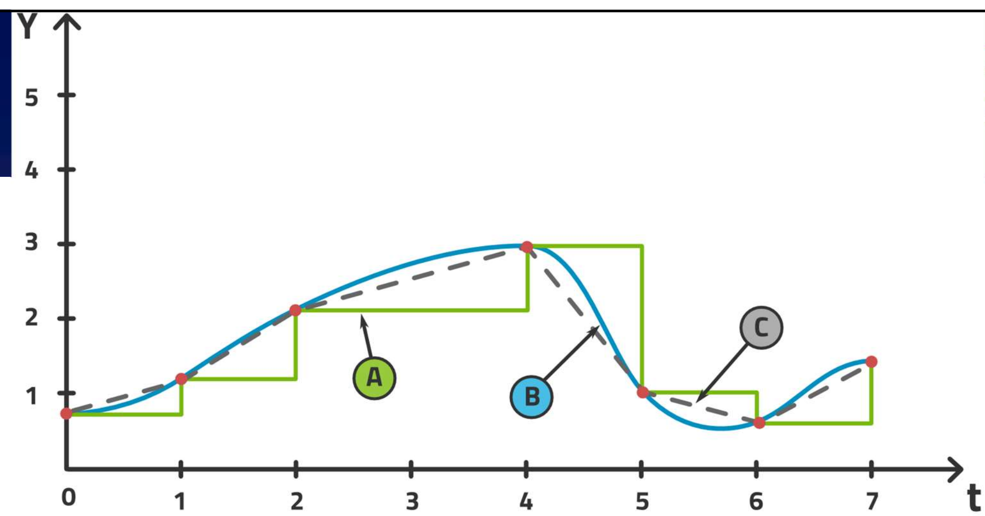

LINEARLY SEPARABLE
NOT LINEARLY SEPARABLE
The domain expert should have some idea of how non-linear the dynamics underlying the dataset is. Different models capture such features better than others.

31



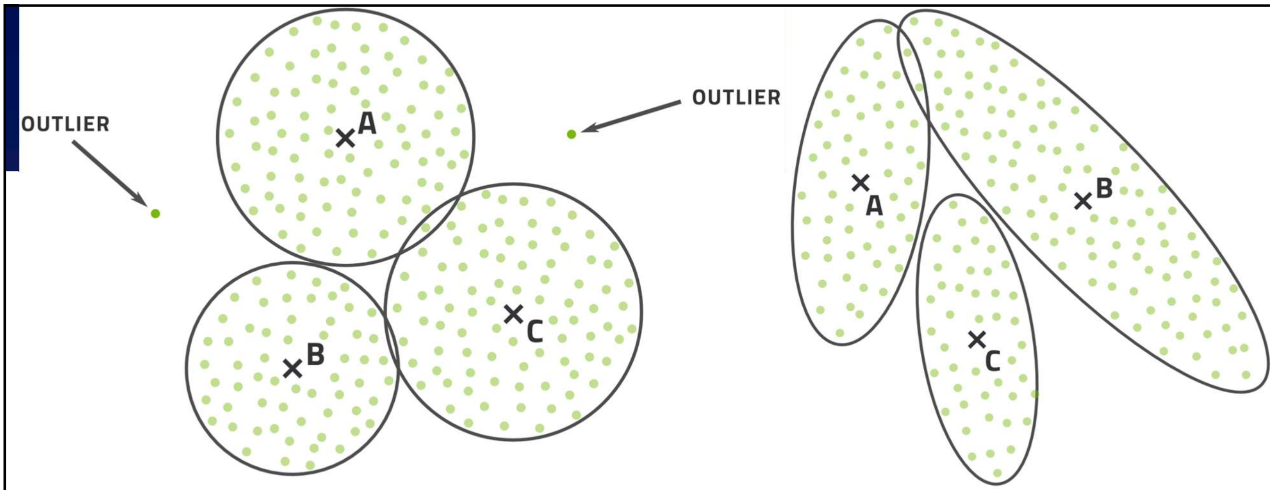
Correlation measures how one variable varies alongside another. The common correlation coefficient is specific to linear correlation.

32



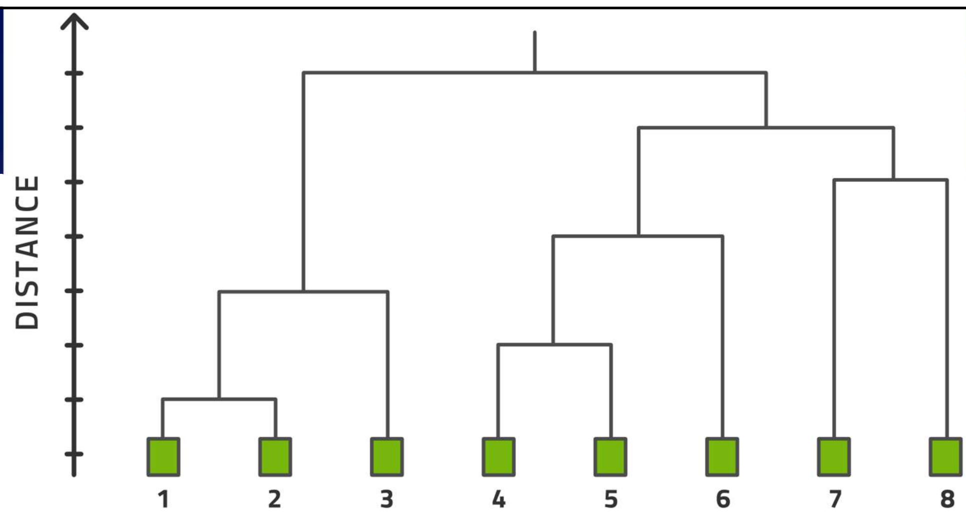
Missing data presents a common challenge. The industry standard way to interpolate is to keep a value constant until the next valid measurement.

33



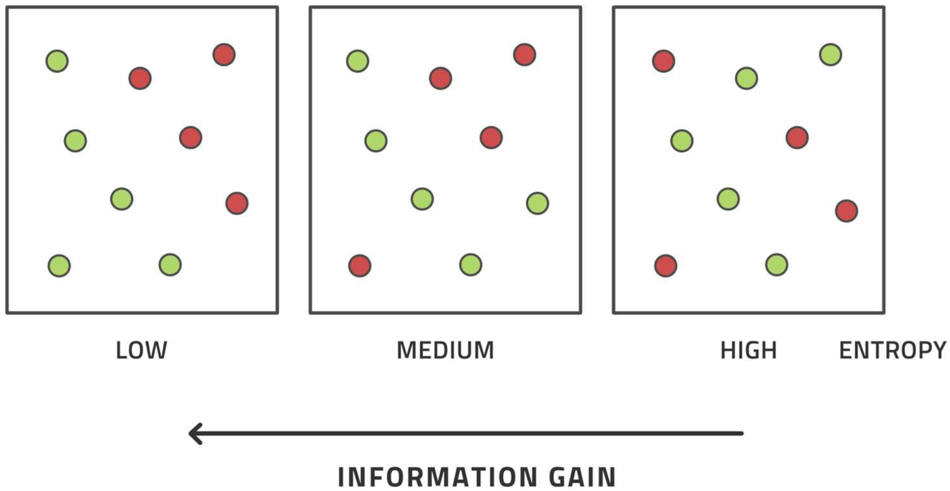
Outliers must be removed to capture the typical behavior. Clustering methods can help to identify such points. Domain knowledge is useful judge what is an outlier and what not.

34



Variables can be grouped into a hierarchy by some similarity measure, e.g. correlation. This can be done top-down or bottom-up.

35

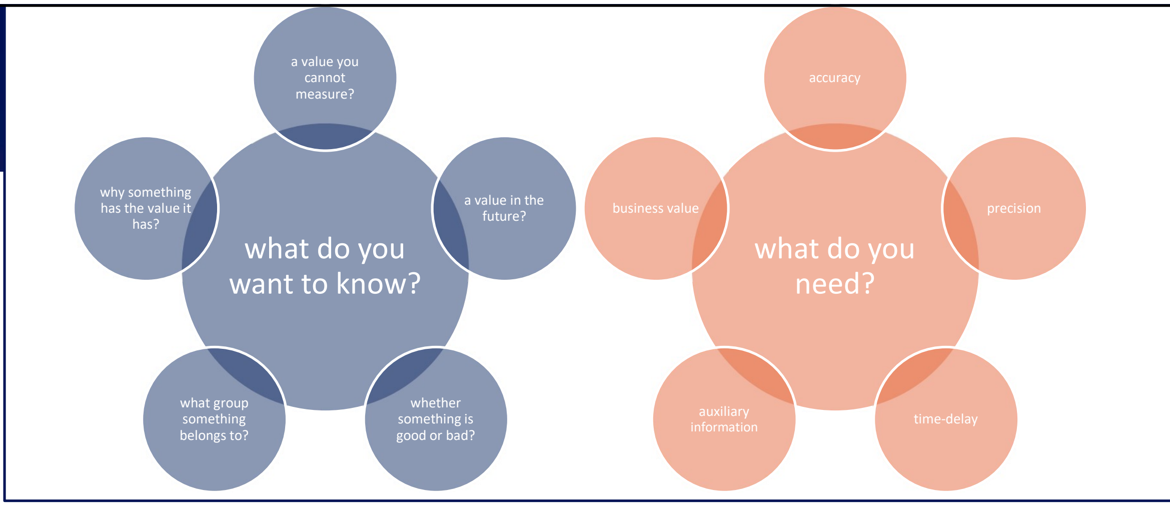


In view of the Bias-Variance-Complexity Trade-Off, we want to reduce the number of variables in our data set to a minimum without sacrificing the information content. Challenge!

36

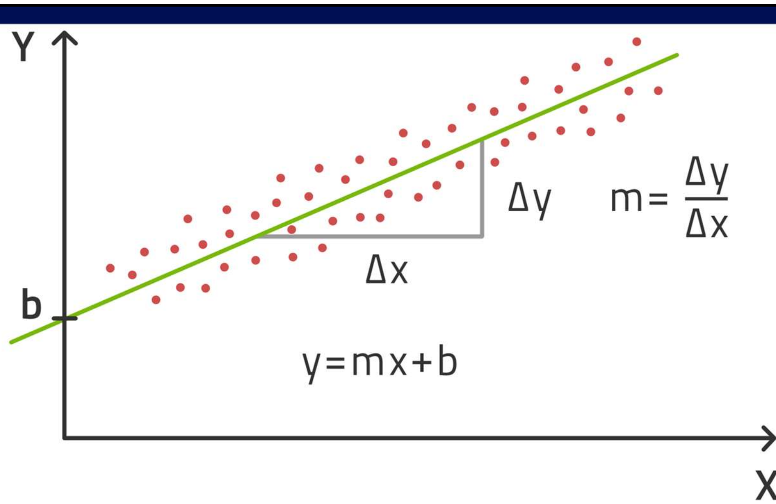
Model Types

37



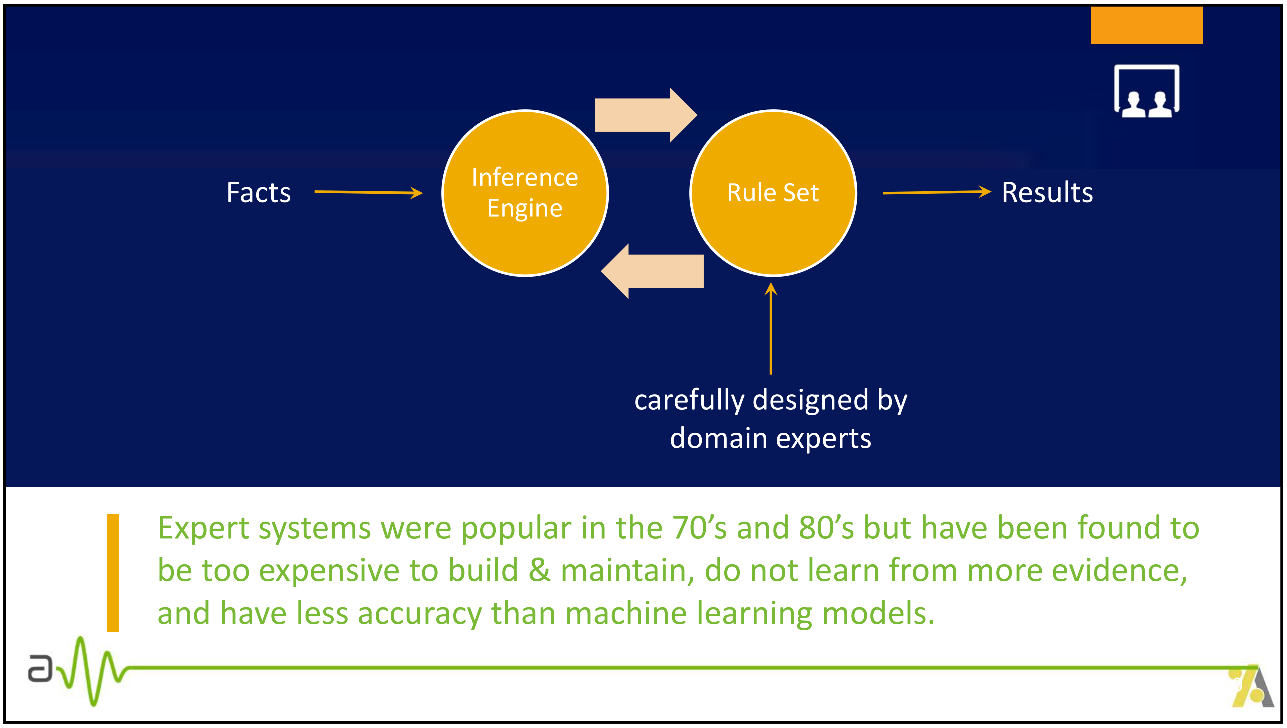
Formulating the question is crucial. Each model and algorithm is suited to a particular kind of question. One cannot put data into machine learning and expect all kinds of novel insights.

38

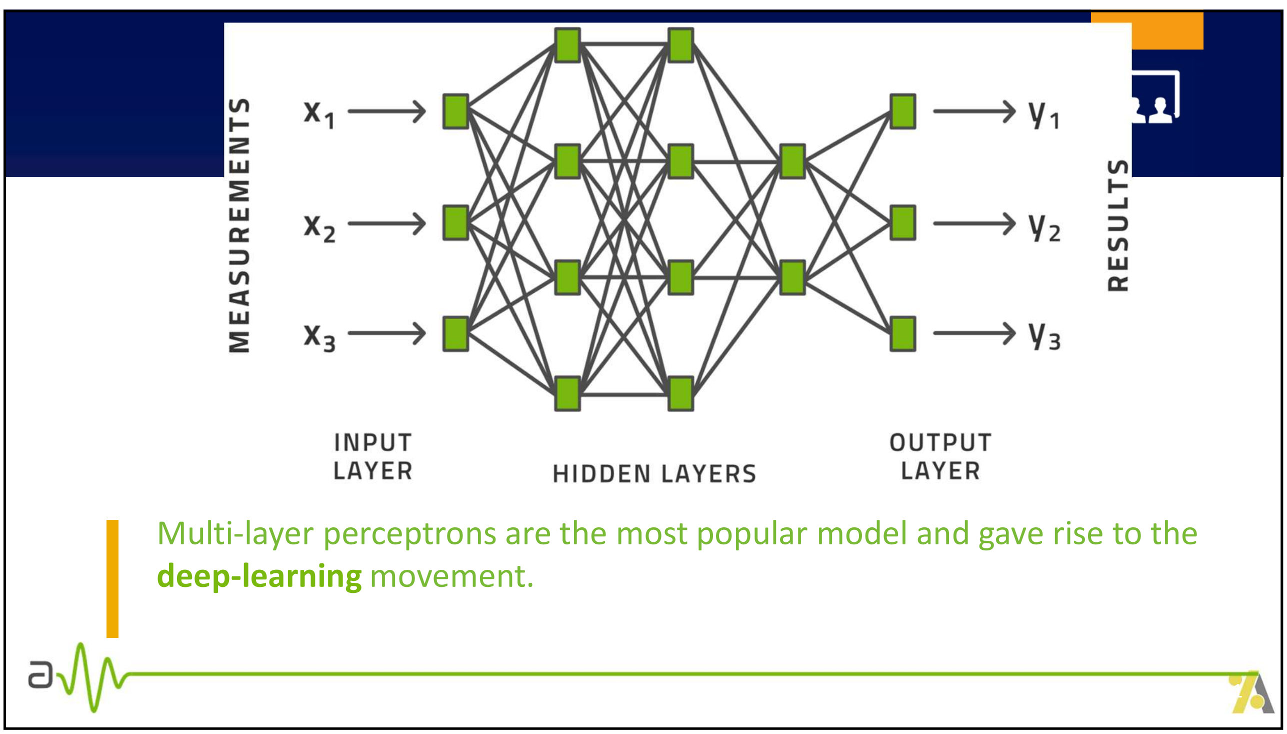


Linear models – possibly with many x's – are useful, at least in a small range of values. These models can be understood by domain experts. However most natural processes are inherently non-linear.

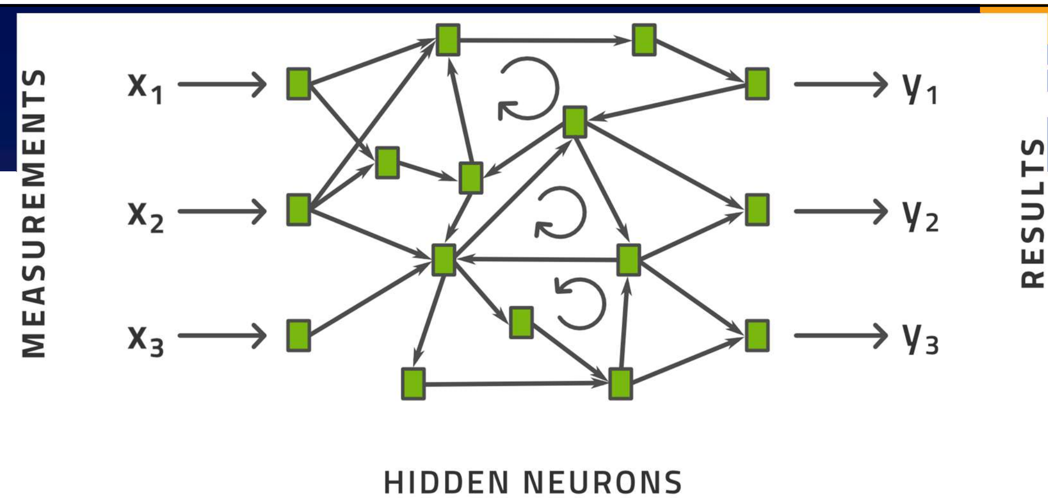
39


Facts
Inference Engine
Rule Set
Results
carefully designed by domain experts
Expert systems were popular in the 70's and 80's but have been found to be too expensive to build & maintain, do not learn from more evidence, and have less accuracy than machine learning models.

40

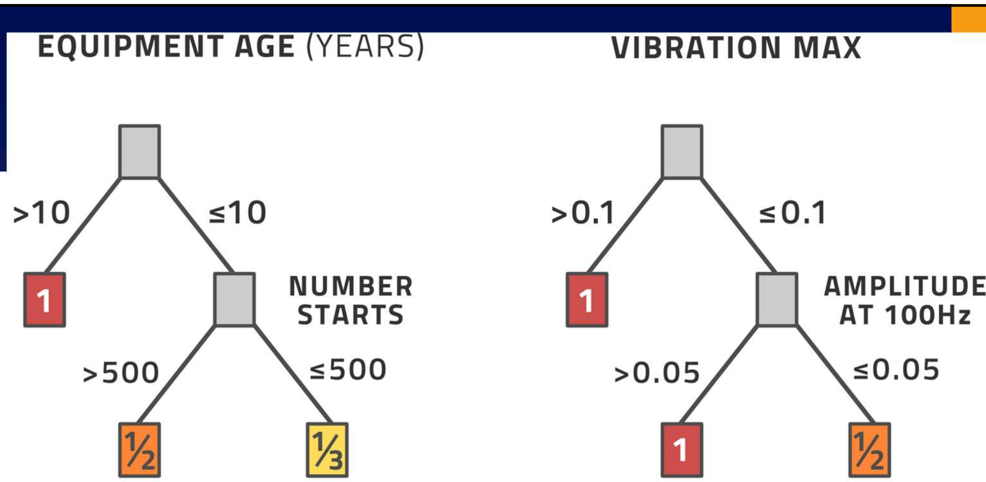

MEASUREMENTS
x_1
x_2
x_3
y_1
y_2
y_3
RESULTS
INPUT LAYER
HIDDEN LAYERS
OUTPUT LAYER
Multi-layer perceptrons are the most popular model and gave rise to the **deep-learning** movement.

41



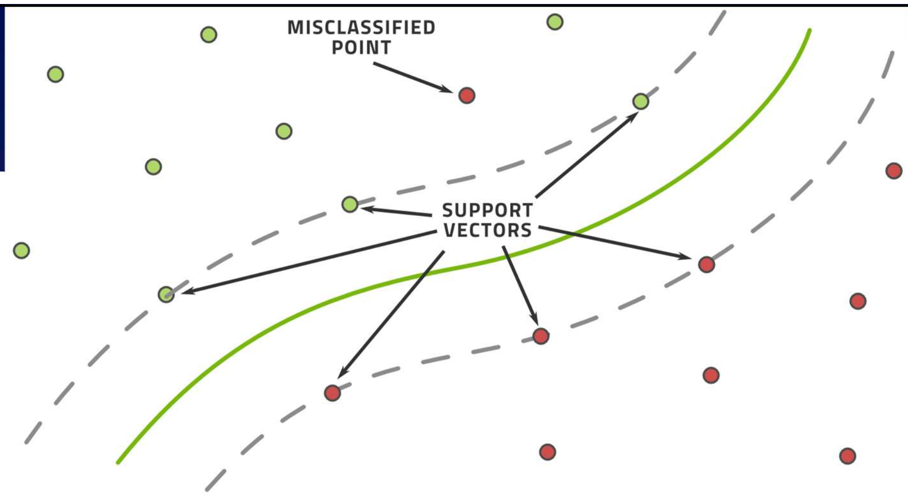
Recurrent neural networks capture the dependence on time that is so crucial for modeling cause-and-effect relationships.

42



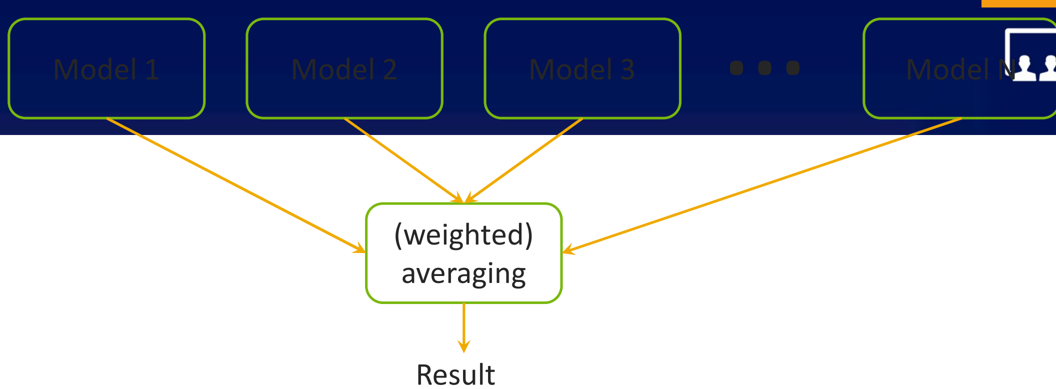
Decision trees or random forests have proven to be the most successful classification models.

43



Support vector machines use a few important training points as special points to define a spline-like boundary between classification regions, possibly in high dimensions.

44



Ensemble models are almost always better than an individual model but need more infrastructure, validation, and care.

45

One Model or Many?
An Example Workflow for Ensemble Modeling

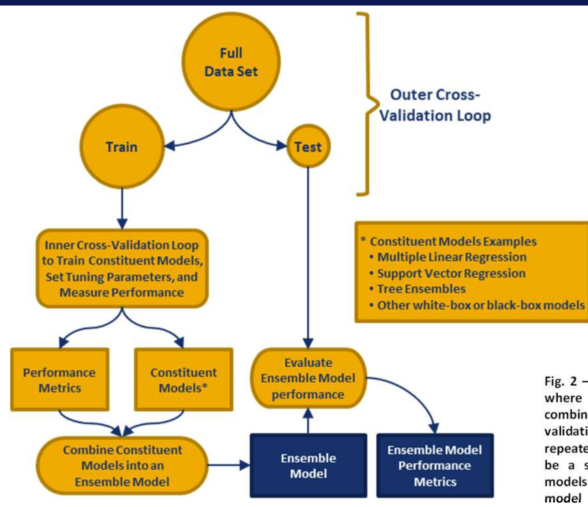
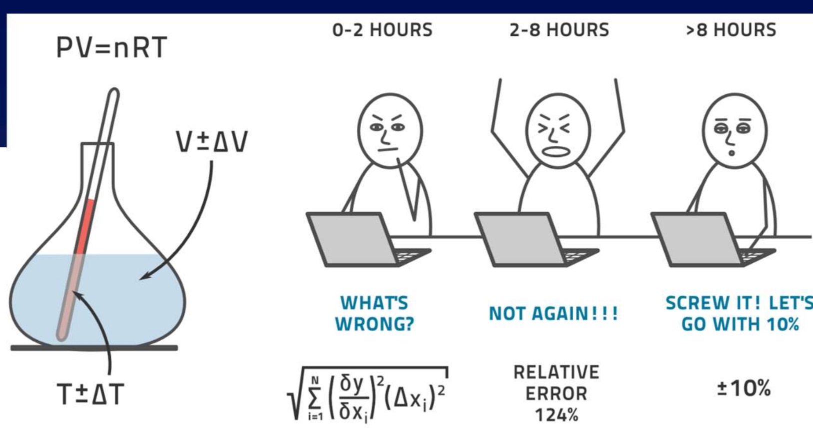


Fig. 2 – An example workflow for ensemble modeling, where constituent models are trained and then combined into a single model (in blue). Nested cross-validation loops are used to control bias and may be repeated to obtain more stable results. Ensembles could be a simple (weighted) average of the constituent models or could use stacking, where the constituent model predictions are treated as feature vector in training the final model.

Source: Mishra *et al*. "Robust Data-Driven Machine-Learning Models for Subsurface Applications: Are We There Yet?." J Pet Technol 73 (2021): 25–30. doi: https://doi.org/10.2118/0321-0025-JPT

46



$$y = f(x)$$

$$\Delta y = \sqrt{\sum_{i=1}^{N}\left(\frac{\partial f(x)}{\partial x_i}\right)^2 (\Delta x_i)^2}$$

$$\Delta P = \frac{nR}{V}\sqrt{(\Delta T)^2 + \left(\frac{T\Delta V}{V}\right)^2}$$

Every measurement is uncertain and so every computation inherits and propagates that uncertainty. We need to know the uncertainty in the model output due to learning and measurement.

47

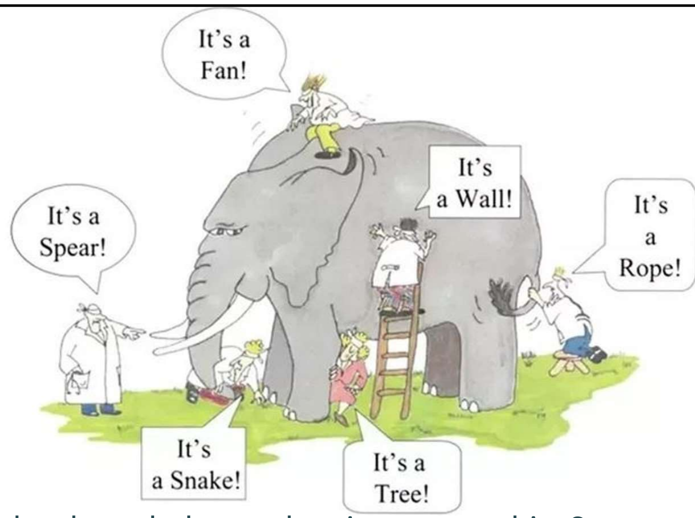
Choosing the Right Model

- What is the desired outcome? classification, regression, time dependency, probability, explanation, visualization.
- Do we have teacher data? Supervised vs. unsupervised. How much data do we have? Bias-variance-complexity.
- Select model that will represent the data in the manner needed. Accuracy will depend mostly on data quality and quantity.
- Do not worry about accuracy at first! This is the last step and mostly differences between 95% and 99% are not felt.

48

Role of Domain Knowledge

49

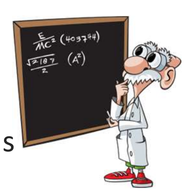


Lack of <u>either</u> knowledge <u>or</u> data is catastrophic. Sensors provide data about all important aspects. Domain knowledge provides context and use-case. Both must be representative and significant.

50

Are the data scientist and domain expert the same person? No. They have different skill sets and knowledge.

data scientist

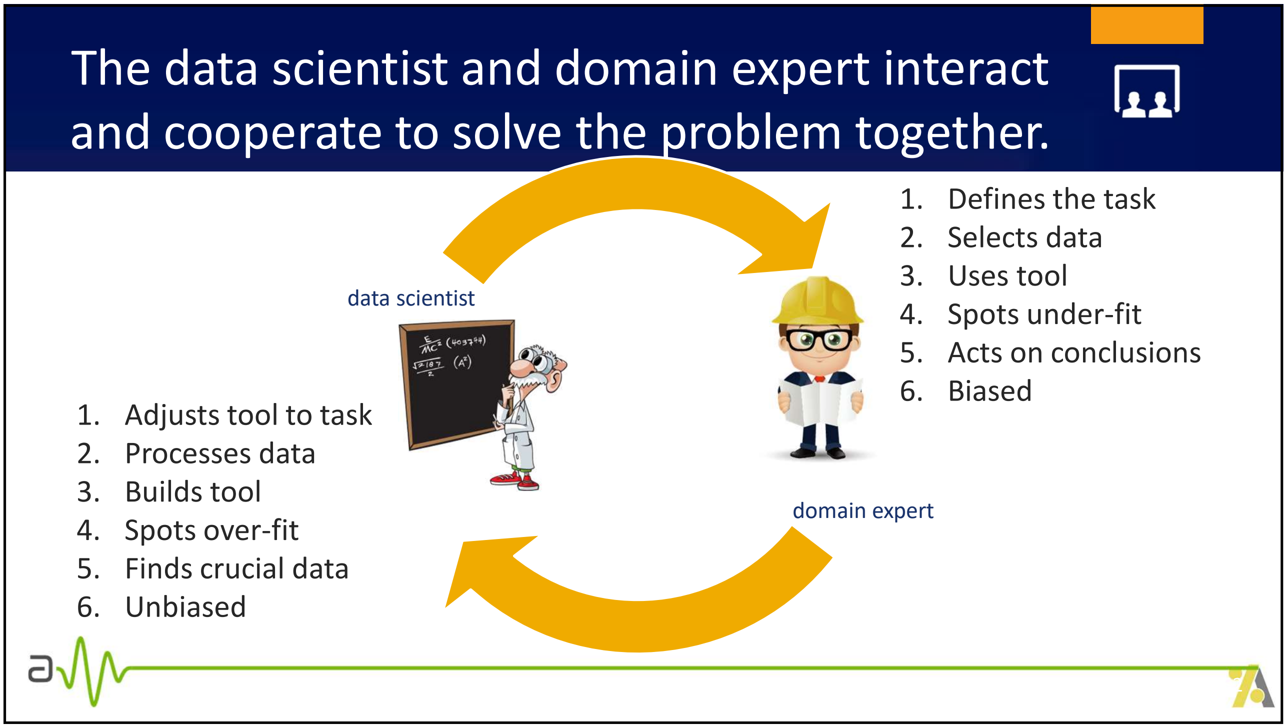


1. Education in data
2. Experience in data
3. Availability of methods
4. Configuration of tools
5. Model quality
6. Communication with technical staff

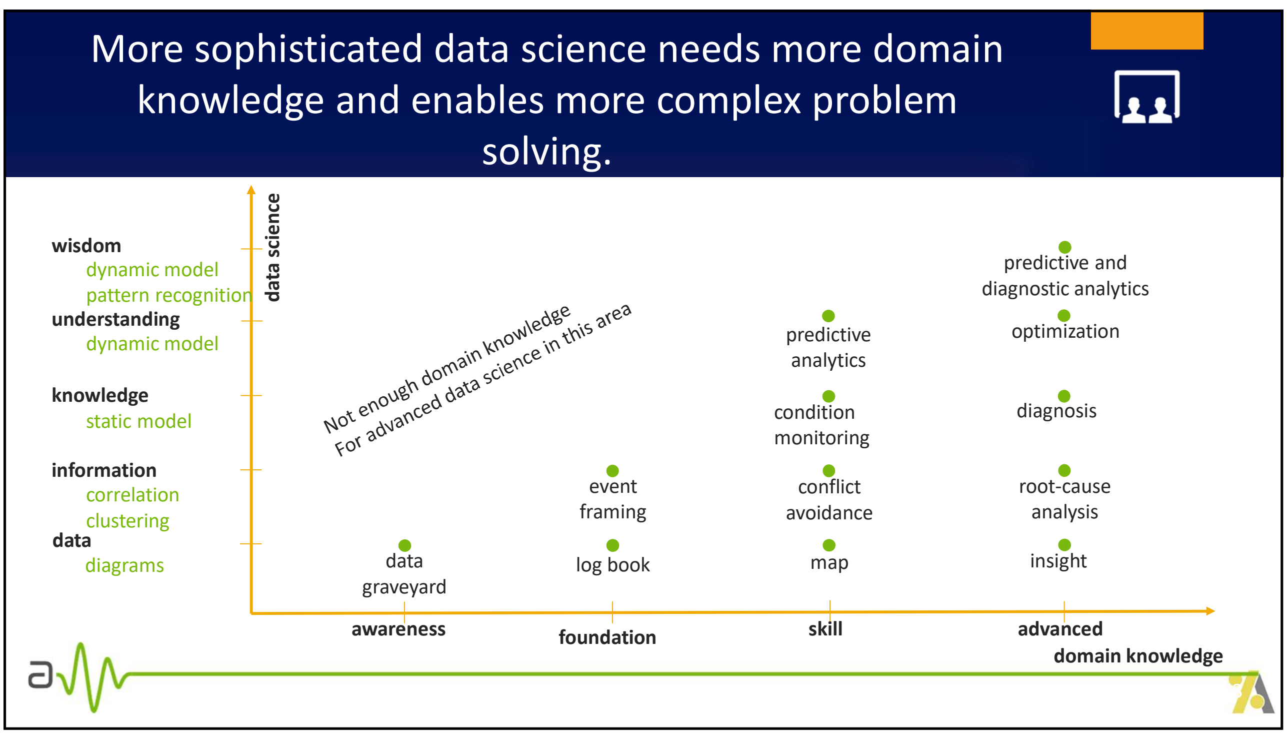
domain expert

1. Education in domain
2. Experience in domain
3. Application of tools
4. Data availability
5. Data quality
6. Communication with intended users

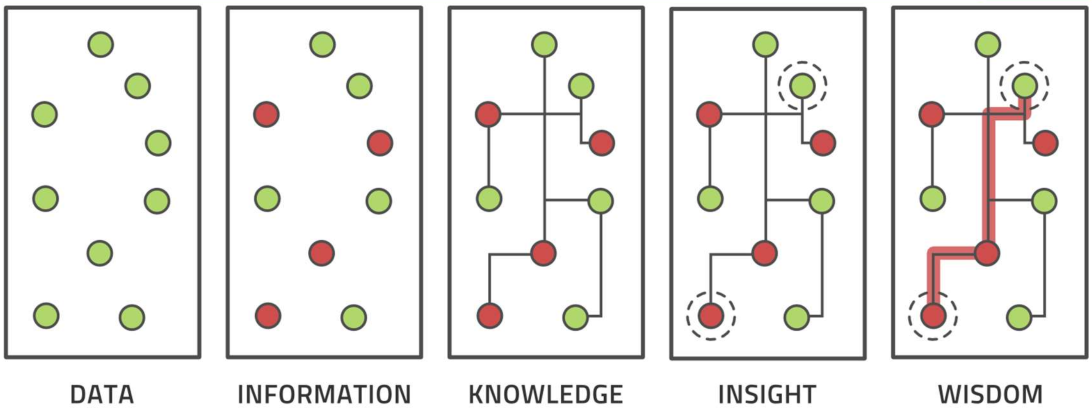
51


The data scientist and domain expert interact and cooperate to solve the problem together.
data scientist
1. Adjusts tool to task
2. Processes data
3. Builds tool
4. Spots over-fit
5. Finds crucial data
6. Unbiased
domain expert
1. Defines the task
2. Selects data
3. Uses tool
4. Spots under-fit
5. Acts on conclusions
6. Biased

52


More sophisticated data science needs more domain knowledge and enables more complex problem solving.
wisdom
dynamic model
pattern recognition
understanding
dynamic model
knowledge
static model
information
correlation
clustering
data
diagrams
data science
Not enough domain knowledge
For advanced data science in this area
data graveyard
event framing
log book
predictive analytics
condition monitoring
conflict avoidance
map
predictive and diagnostic analytics
optimization
diagnosis
root-cause analysis
insight
awareness
foundation
skill
advanced
domain knowledge

53

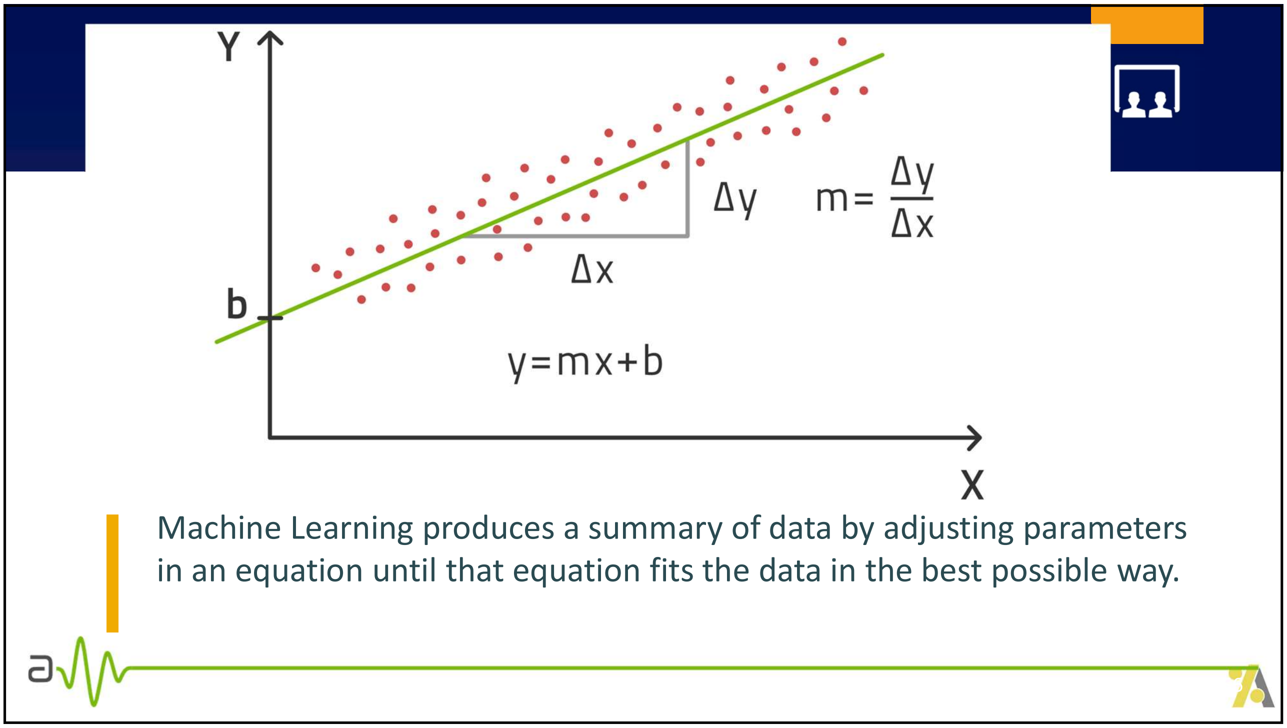


We gradually progress from data to wisdom. It is at information and knowledge that we need domain expert input. Insight and wisdom are then outcomes provided by machine learning on the combined input.

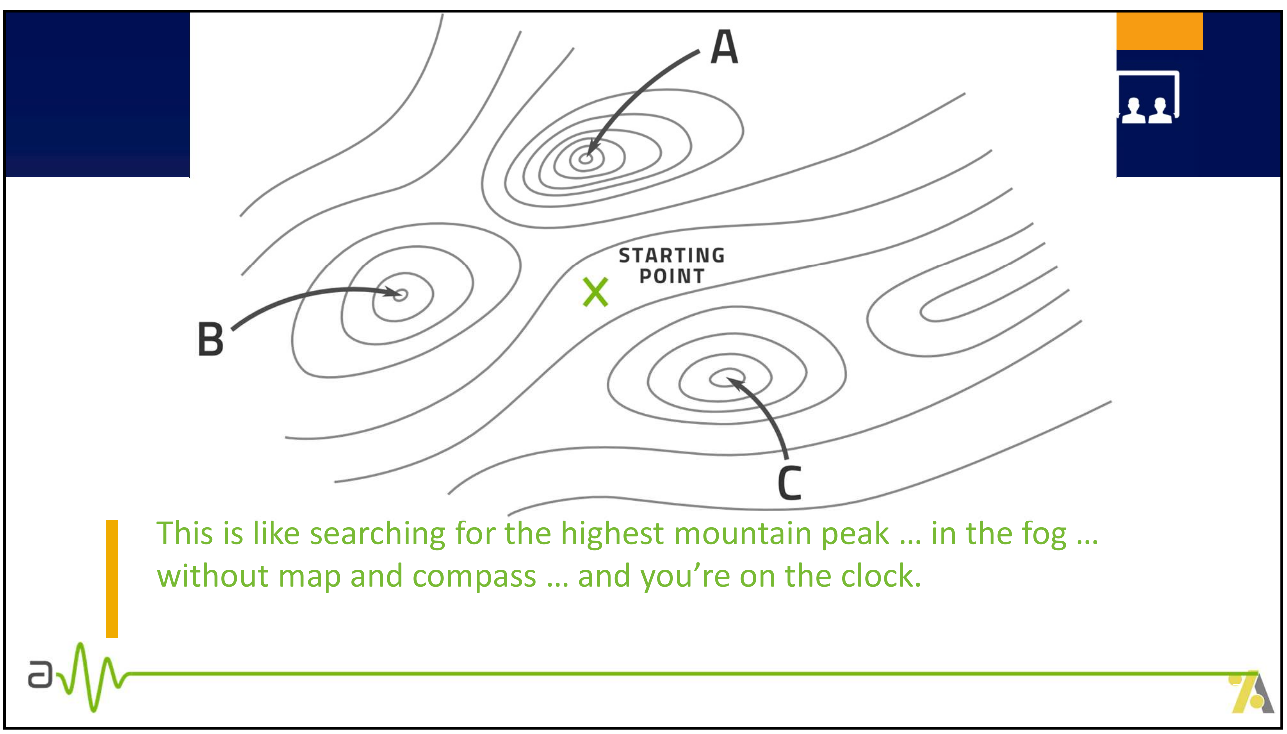
54

Training a Model

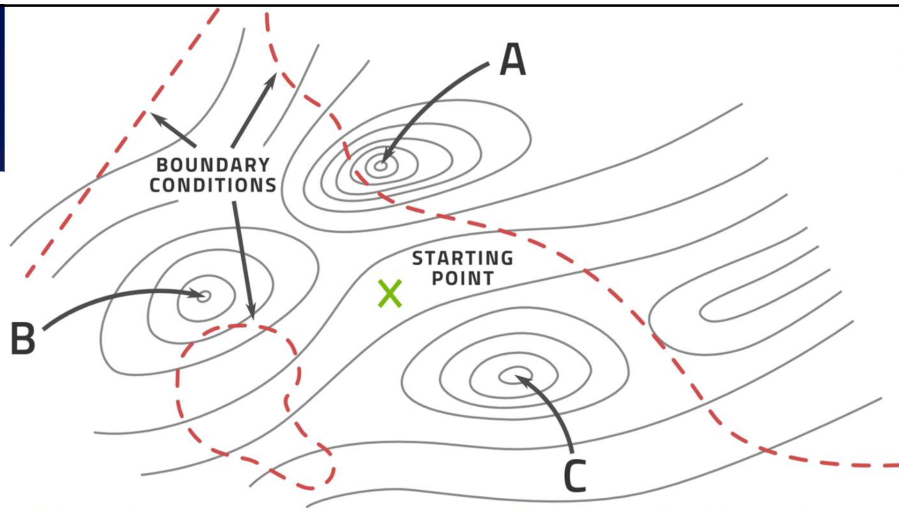
55


Y
b
X
Δy
Δx
m= Δy/Δx
y=mx+b
Machine Learning produces a summary of data by adjusting parameters in an equation until that equation fits the data in the best possible way.

56

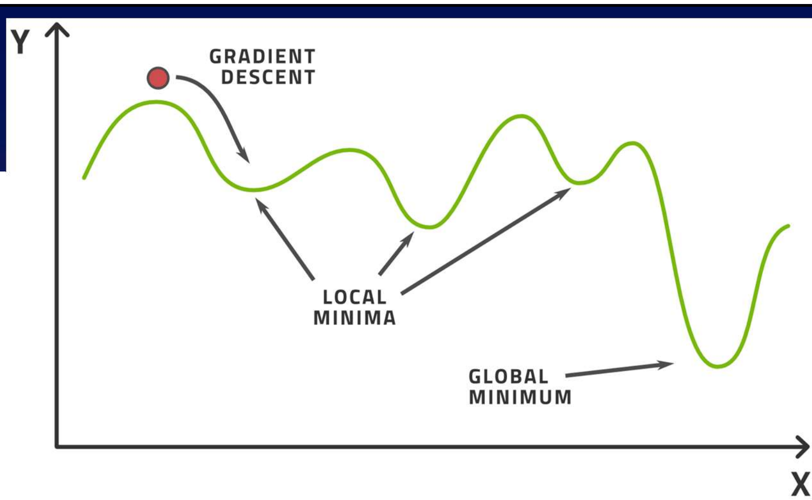

A
B
C
STARTING POINT
This is like searching for the highest mountain peak … in the fog … without map and compass … and you're on the clock.

57



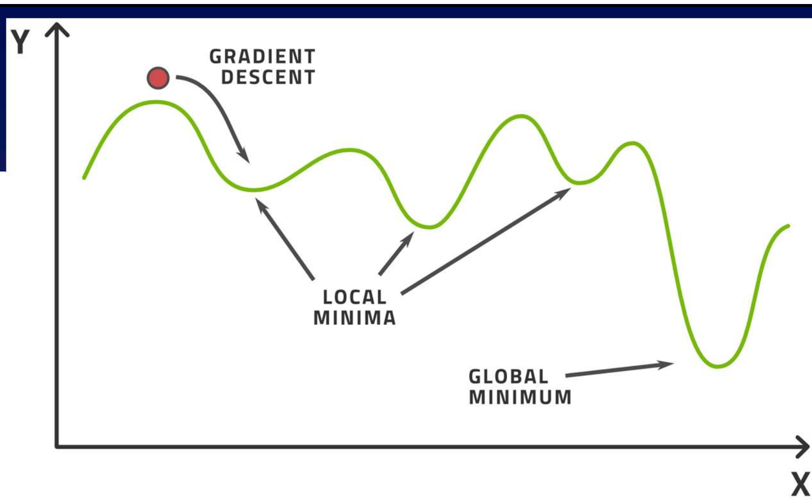
Real life tasks have many restrictions that are called boundary conditions in mathematics: Safety requirements, process conditions, equipment limitations, etc. These must be included.

58



Local minima is good but the global minimum is best. Gradient descent is by far the most common approach. Due to getting stuck in local minima, we typically do several random restarts.

59

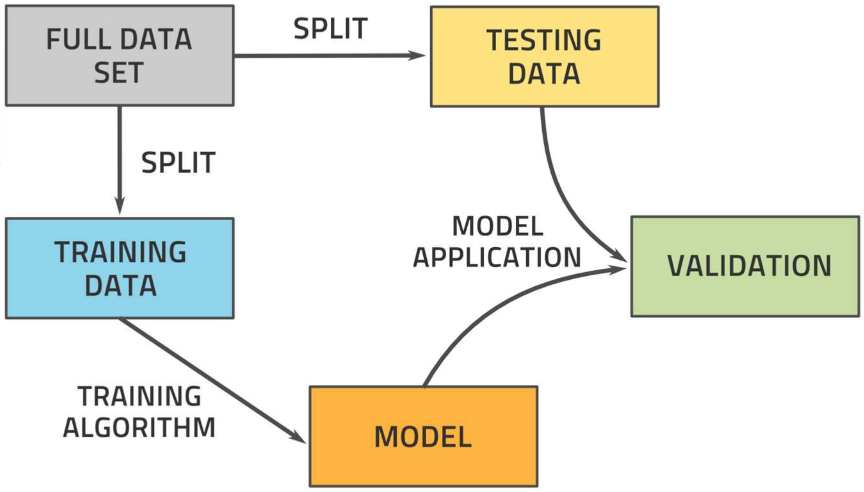


General principles, like gradient-descent, must be realized for specific model types => **learning algorithm**. These have parameters that must be **tuned** either by another algorithm or by expert knowledge.

60

How good is my model?

61



We must test the model on data that has not been used to train the model. The full data set must be split into two parts. Both must be significant and representative, and they must be similar to each other.

62

Regression

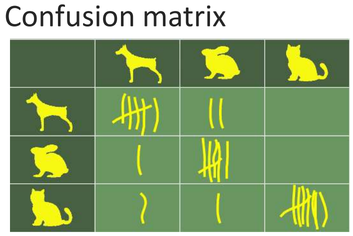
$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(y_i - y'_i)^2}{N}}$$

$$MAE = \sum_{i=1}^{N} \frac{|y_i - y'_i|}{N}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - y'_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2}$$

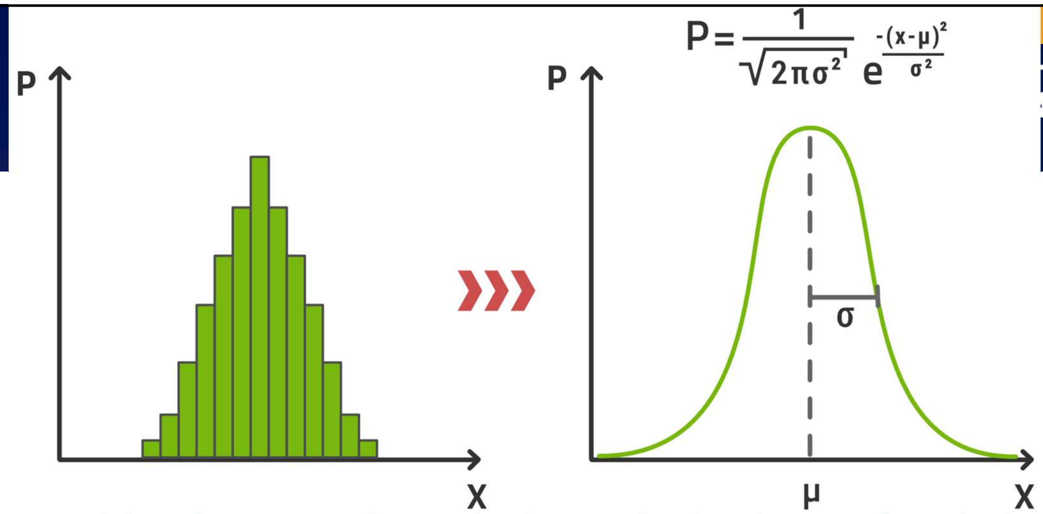
Classification

- Accuracy
- False positive rate
- False negative rate
- Statistical hypothesis testing
- Confusion matrix



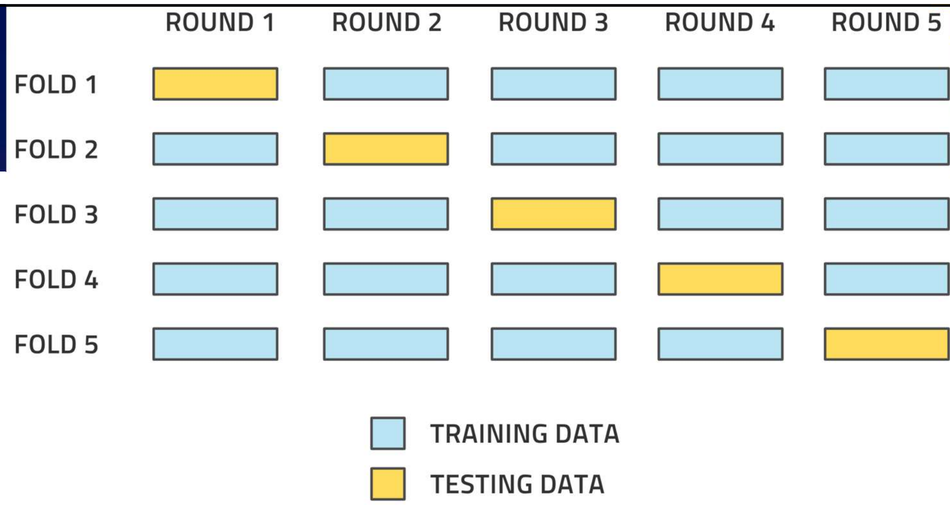
Various methods to quantify model performance exist. They must be interpreted carefully. Unless the model is excellent, it is hard to quantify its goodness using a single number.

63



Model performance is best judged using the distribution of residuals (model - measurement) where bias and variance can be seen readily.

64



In order to assess the bias and variance of the model, we perform cross-validation: We break the full data set into several training and testing data sets.

65

$$bias = \sum_{i=1}^{N} bias_i$$

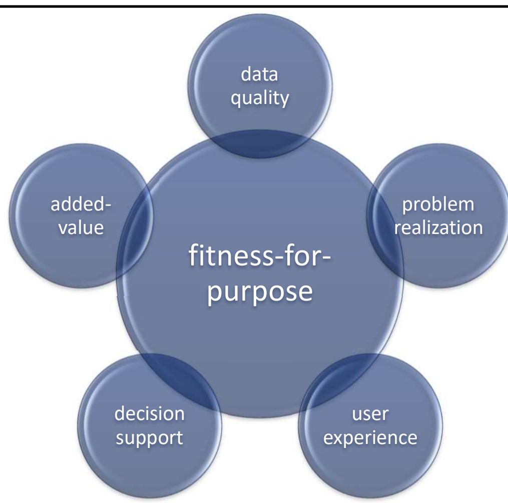
$$\sigma = \sqrt{\frac{bias(1-bias)}{N}}$$

$$error = bias \pm z_L\sigma$$

Confidence Level	z_L
80%	1.28
90%	1.645
95%	1.960
98%	2.33
99%	2.576
99.5%	2.807
99.9%	3.291

In cross-correlation, we have multiple models. If we take enough rounds (> 30), then these values are normally distributed (central limit theorem). What is the true error w.r.t. the population?

66



Goodness-of-fit measures mathematical accuracy. We must also determine fitness-for-purpose: Does this model deliver the added value we are seeking in the business problem?

67

Introduction to Toolsets

68

Programming Languages

R

- Scripting for analysis
- More visualization
- More academic
- Higher level control
- Broader functionality

Python

- Full programming language
- Less visualization
- Less academic
- Lower level control
- Narrower Functionality

- Large ecosystem (functions, people)
- Requires expert user
- Useful for prototyping – Product development in separate environment

69

Frameworks

- Frameworks are high-level languages that do not need full programming
- They offer good interfaces to a variety of algorithms
- Generally useful only for model building and analysis
- Data preparation, cleaning, import/export, interactivity, productive use is done elsewhere

- scikit-learn
- TensorFlow
- PyTorch
- WEKA
- Hadoop/spark

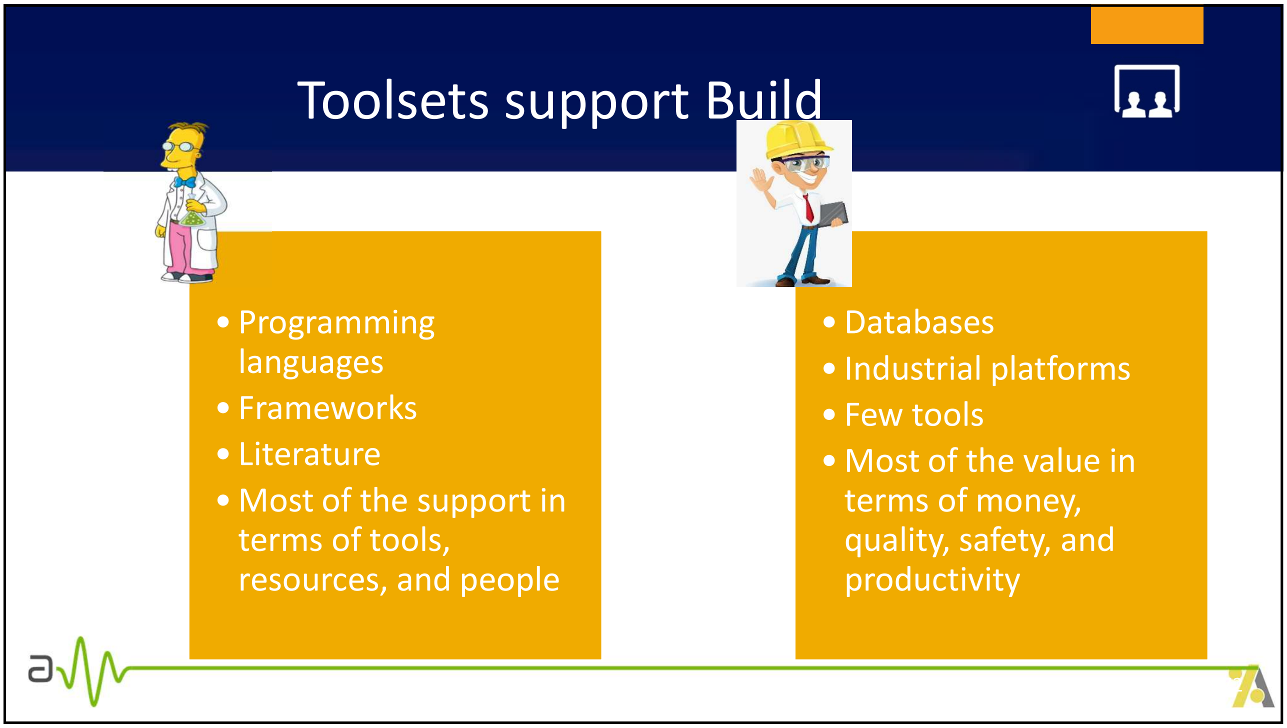
70

Deploying is More Complex than Training

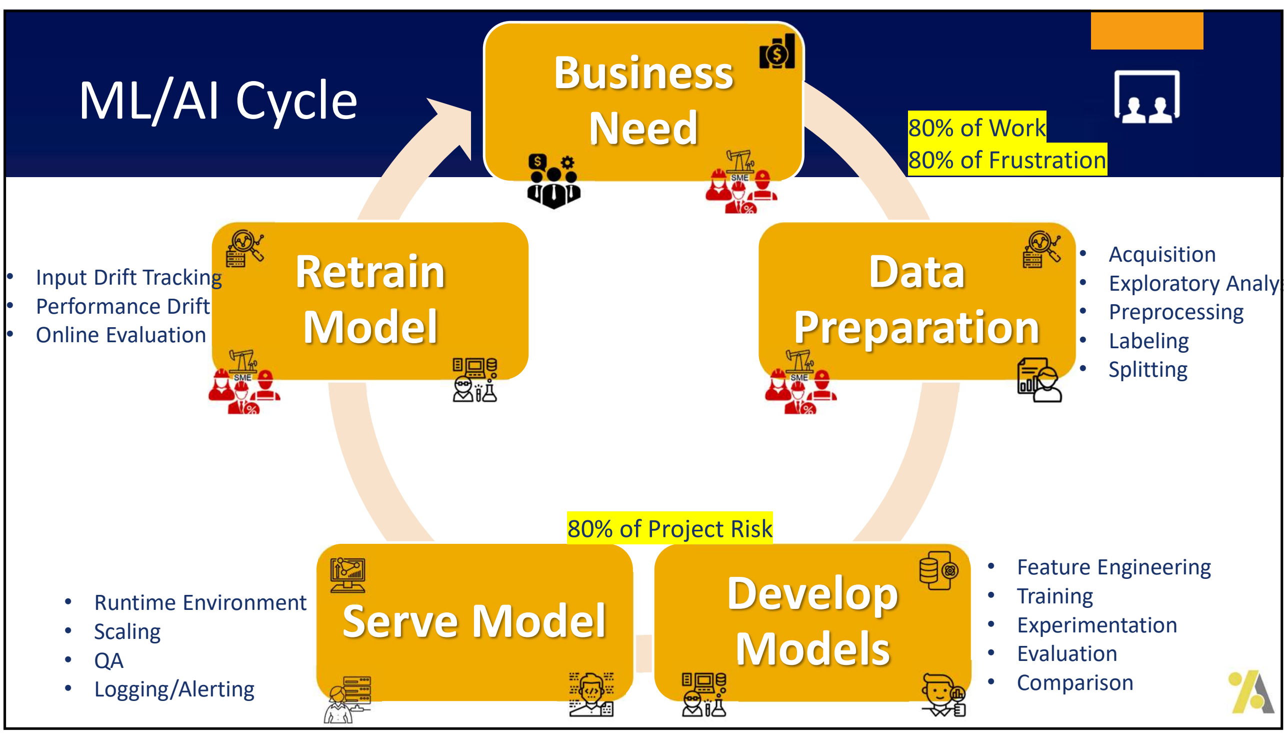
- Data Cleansing
- Analysis
- Learning
- Assessment
- Fine-Tuning
- Data Science

- Data Interfaces
- User Interface
- Scalability (speed, space)
- Visualization
- Workflow Integration
- Management & Reporting

71


Toolsets support Build
• Programming languages
• Frameworks
• Literature
• Most of the support in terms of tools, resources, and people
• Databases
• Industrial platforms
• Few tools
• Most of the value in terms of money, quality, safety, and productivity

72


ML/AI Cycle
Business Need
80% of Work
80% of Frustration
Data Preparation
• Acquisition
• Exploratory Analy
• Preprocessing
• Labeling
• Splitting
Retrain Model
• Input Drift Tracking
• Performance Drift
• Online Evaluation
80% of Project Risk
Serve Model
• Runtime Environment
• Scaling
• QA
• Logging/Alerting
Develop Models
• Feature Engineering
• Training
• Experimentation
• Evaluation
• Comparison

73

Plentiful Online Resources on Developing ML Skills

1. Free online machine learning curriculum (huyenchip.com): https://huyenchip.com/2019/08/05/free-online-machine-learning-curriculum.html
2. 2020 Machine Learning Roadmap (95% valid for 2023): https://www.youtube.com/watch?v=pHiMN_gy9mk
3. …
4. …

74

75

All Jokes aside, how do we begin with AI/ML in ARTIFICIAL-Lift?

76

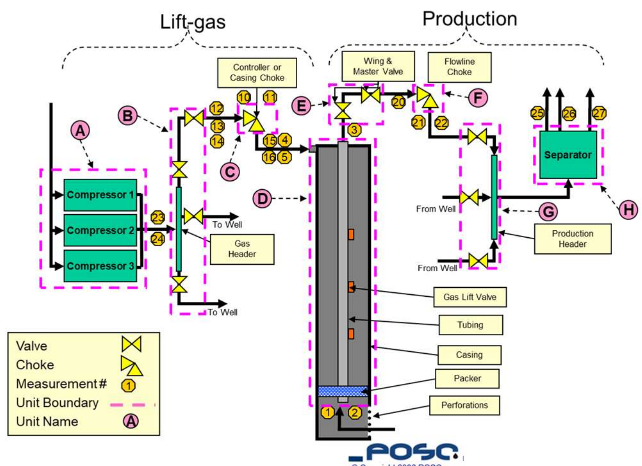
Identifying Gas-Lift Use Cases

- Below problems encountered in gas lift and discuss which category of ML (supervised or unsupervised) they might belong to.
  1. Slugging
  2. Optimal injection depth
  3. Over/under-injection
  4. Multi-pointing

77

Identifying Gas-Lift Use Cases

Draw Boundary around items with no measurements



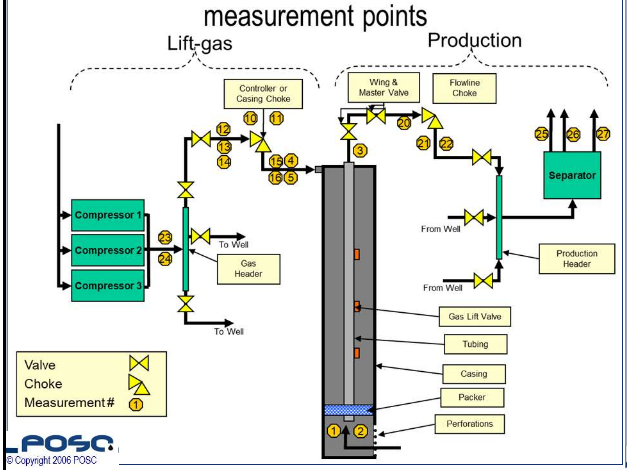
Control volumes around critical measurement points

Source: http://w3.energistics.org/schema/prodml_v1.2.0_data/doc/PRODML_Gas_Lift_Scenario_Network_Construction.ppt

78

Identifying Gas-Lift Use Cases

Draw the real-world diagram indicating measurement points



Description of Measurement Points in Diagram

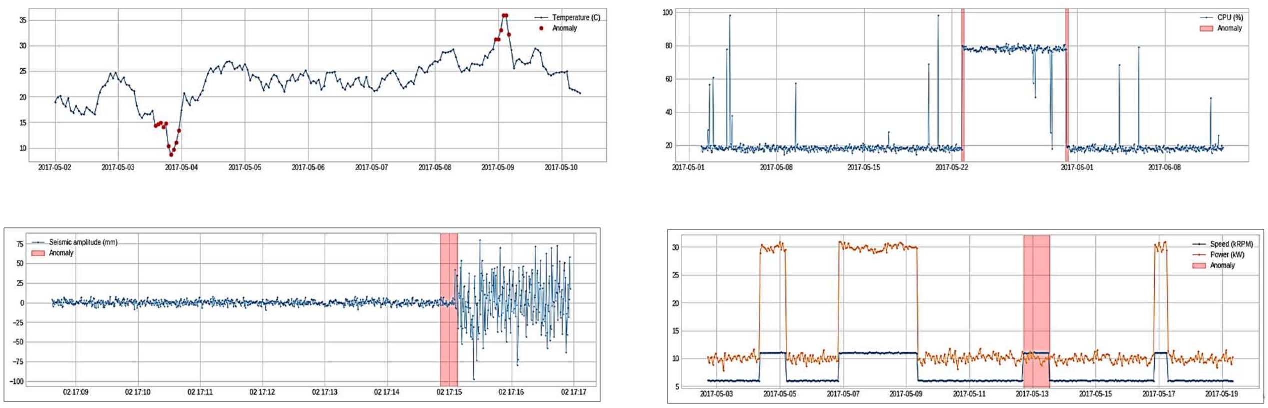
1	Bottom hole pressure	Well (Subsurface)
2	Bottom hole temperature	Well (Subsurface)
3	Flowing tubing head pressure	Well (Surface)
4	Annulus pressure	Well (Surface)
5	Lift gas pressure (CHP)	Well (Surface)
6	SSSV status	Well (Surface)
7	SSV status	Well (Surface)
8	Chemical injection flow (Injection gas)	Well (Surface)
9	Chemical injection (Injection gas) ppm	N/A
10	Lift gas control valve percent open	Well (Surface)
11	Lift gas rate setpoint	N/A
12	Lift gas pressure (upstream of meter)	Well (Surface)
13	Lift gas meter differential pressure (dp)	Well (Surface)
14	Lift gas meter temperature (upstream of control valve)	Well (Surface)
15	Lift gas flow rate	N/A
16	Lift gas temperature	n/a
17	Chemical injection flow (Production)	Well (Surface)
18	Chemical injection (Production) ppm	N/A
19	Sand detection	Well (Flow line)
20	Flow line pressure differentual	Well (Flow line)
21	Flow line pressure	Well (Flow line)
22	Flow line temperature	Well (Flow line)
23	Bulk lift gas temperature	Injection Manifold
24	Bulk lift gas pressure	Injection Manifold
25	Total gas production rate	N/A
26	Water production rate	N/A
27	Net oil production rate	N/A

Source: http://w3.energistics.org/schema/prodml_v1.2.0_data/doc/PRODML_Gas_Lift_Scenario_Network_Construction.ppt

79

Anomaly Detection

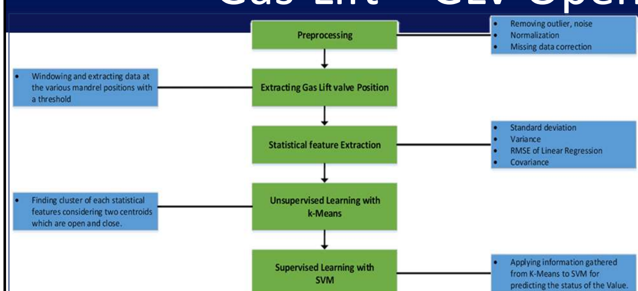
Objective:
Discover deviation in data potentially signaling an event like process upset, imminent failure, defect.
Learn what "normal" data looks like then use data to detect abnormal instances

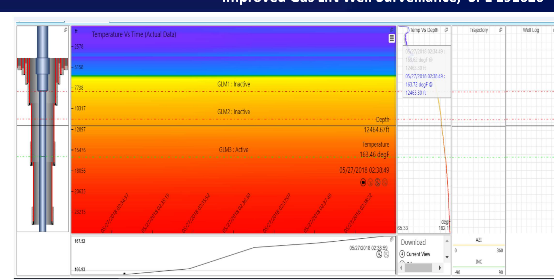


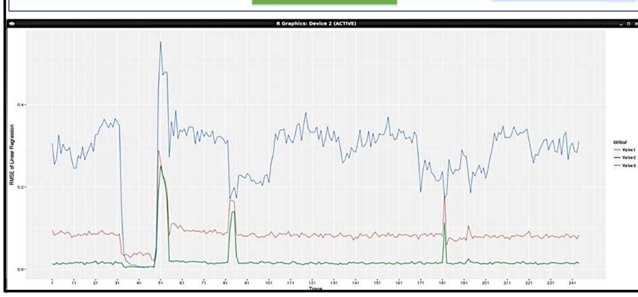
80

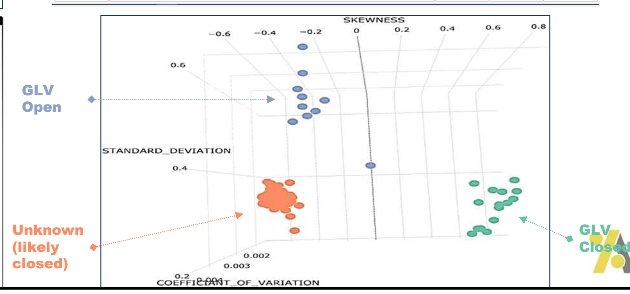
DTS Analytics for Anomaly Detection in Gas-Lift – GLV Open or Close??

Source: Bello *et al.* (2018) Integrating Downhole Temperature Sensing Datasets and Visual Analytics for Improved Gas Lift Well Surveillance, SPE 191626

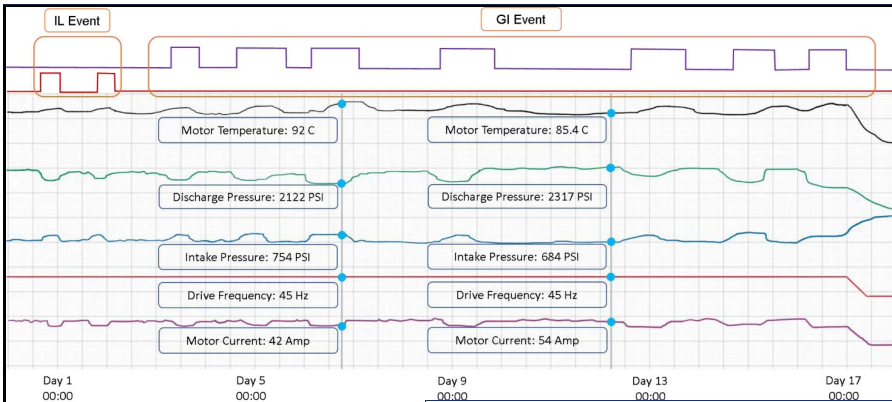


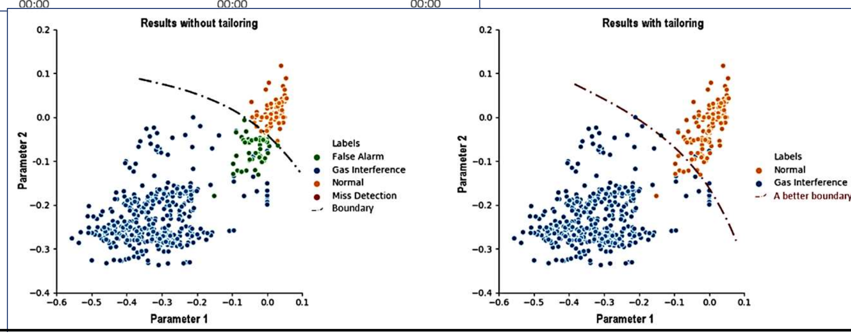






81





Source: Nguyen *et al.* (2020), Real-Time Automated Event Detection Framework for Electrical Submersible Pumps, SPE 203277

82

Failure Prediction
Machine Learning approach for ESP

- ESP Well and Component Failure Prediction in Advance using Engineered Analytics - A Breakthrough in Minimizing Unscheduled Subsurface Deferments
  - Introduction
  - Analytical Data Modeling
    - Machine Learning
    - Approach and Process
  - Data Preparation
    - Data Gathering
    - Data Pre-Processing
  - Model Building
    - Dataset Sample Selection
    - Model Training
    - Feature Selection and Engineering
    - Parameter Tuning
  - Model Testing
    - Creating a Dataset for Model Testing
    - Model Scoring
    - Test Results
  - Model Evaluation by Well Owners (Operator's Blind Test)
    - Methodology
    - Operator's Blind Test Results
  - Value Generated by the Analytical Model
  - Conclusions and the way ahead
- References

Static Data

Installation and Commissioning data

Well Completion information

ESP Well Failure history

Ad-hoc information (emails, spreadsheet, reports, photos)

Dynamic Data

Downhole Gauge: Pump Discharge & Intake pressures, Motor temperature, Intake temperature, vibration etc.

Motor Controller: Electrical readings, well running status and trip conditions.

Surface Pressure Gauges: Tubing head pressure, Flow line pressure

Calculated Data

Well Model Based Calculated Data: Inferred production and Pump Efficiencies calculated from the Physics based models in Real Time

Supervised learning approach: Random Forest based classifier developed using above and engineered features like standard deviations, param differences across fixed time intervals, differential pressures, etc.

Well Level Failure and component level failure models developed.

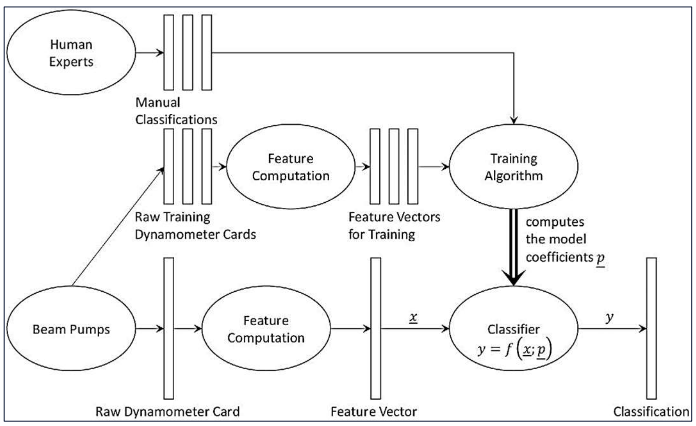
Source: Marin *et al.* (2019) ESP Well and Component Failure Prediction in Advance using Engineered Analytics - A Breakthrough in Minimizing Unscheduled Subsurface Deferments, SPE-197806

83

Predictive Maintenance for Rod Pumps

Source: Bangert & Sharaf: SPE 195295 – "Predictive Maintenance for Rod Pumps", 2019

- **Procedure of generating and using classification model**
- **Formula taking card's features as input and produces a class number as output**



35292 cards from 299 beam pumps from Bahrain	
Normal	
Fluid Pound (Slight)	Inoperative Pump Hitting Down
Fluid Pound (Severe)	TV or Plunger Leak
Pump Hitting Down	SV, TV Leak or Gas Interference
Pump Hitting Up & Down	Inoperative Pump

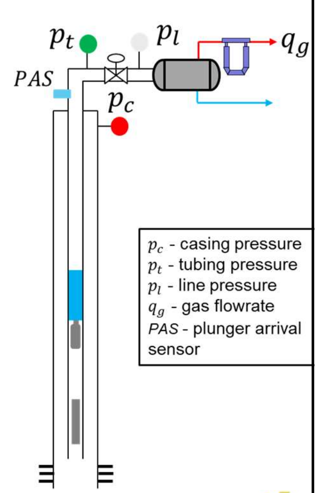
Results:
Using stochastic gradient-boosted decision tree model
Accuracy = 99.9%   Evenly distributed errors

84

Virtual Flow Meter – Plunger Lift

Source: Akhiiratdinov *et al.* (2020) Data Analytics Application for Conventional Plunger Lift Modeling and Optimization, SPE 201144

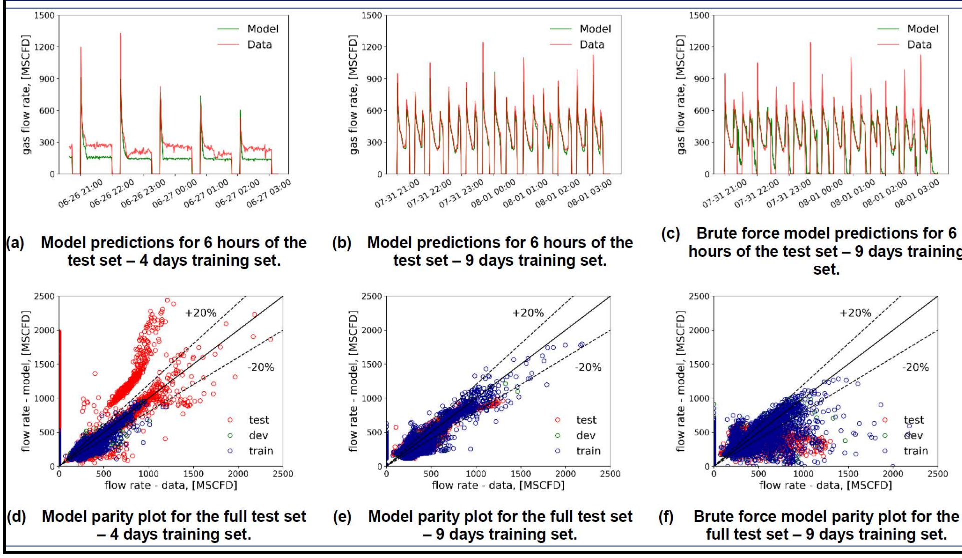
- Using plunger arrival and pressures data, determine the gas flow rate to back-allocate the flow rates.
  - Challenging or computationally expensive using traditional physics-based model
- Approach:
  - ANN – 3 hidden layers, 10 nodes each layer
    - Brute force approach utilizing raw pressure and flow rate data in the ANN does not yield acceptable results
    - Physics-guided features – pressure differentials and flow rate transformations – provide better model performance.
- More training data helps evolve model into better prediction performance.



85

Virtual Flow Meter – Plunger Lift

Source: Akhiiratdinov *et al.* (2020) Data Analytics Application for Conventional Plunger Lift Modeling and Optimization, SPE 201144



(a) Model predictions for 6 hours of the test set – 4 days training set.

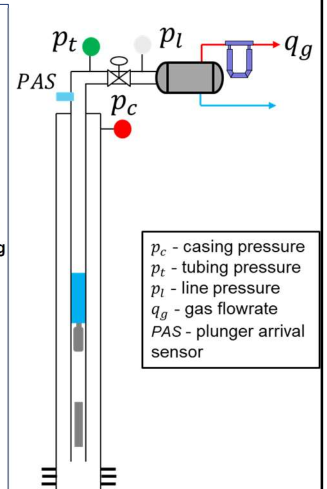
(b) Model predictions for 6 hours of the test set – 9 days training set.

(c) Brute force model predictions for 6 hours of the test set – 9 days training set.

(d) Model parity plot for the full test set – 4 days training set.

(e) Model parity plot for the full test set – 9 days training set.

(f) Brute force model parity plot for the full test set – 9 days training set.



86

Questions & Discussions

87

Pan American ENERGY

2.1. Setup Environment for Working with Hands-on Examples

Dr. Rajan Chokshi

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

Outline

- How do we develop & work with ML/AI examples in this class?
  - Full-fledged Development Setup
    - Python Editor, Interpreter, Package libraries.
    - Not undoable but can become complex and frustrating for classroom purpose.
  - Keep everything in cloud and access it from a browser.
  - Need a constant internet connection though.
- Google Colab
- GitHub Repository

2

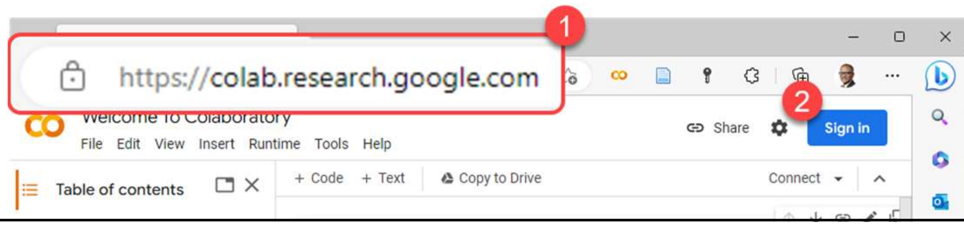
Google Colab

- Google Collaboratory
  - Cloud based & free with a Google account (which is also free).
  - Need only browser compatible with Colab.
    - Chrome or Edge works well
  - Provides editors and interpreters for popular programming languages
- Github Repository
  - Provides storage for scripts and data files.

3

Setup

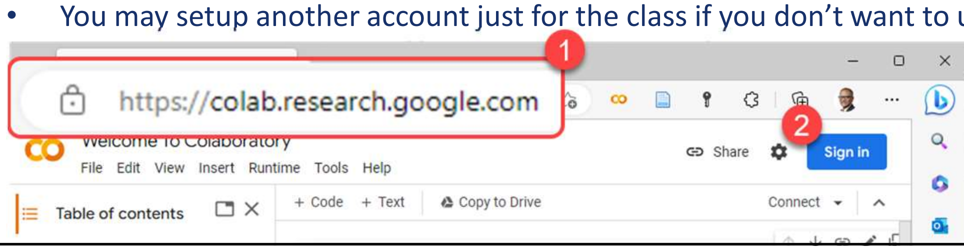
1. Open a browser (Chrome preferred but will work with Edge)
   - Go to https://colab.research.google.com/
2. Sign in into your Google account by clicking on the right top button.
   - You may setup another account just for the class if you don't want to use your account. Go to https://accounts.google.com/signup



4

Setup

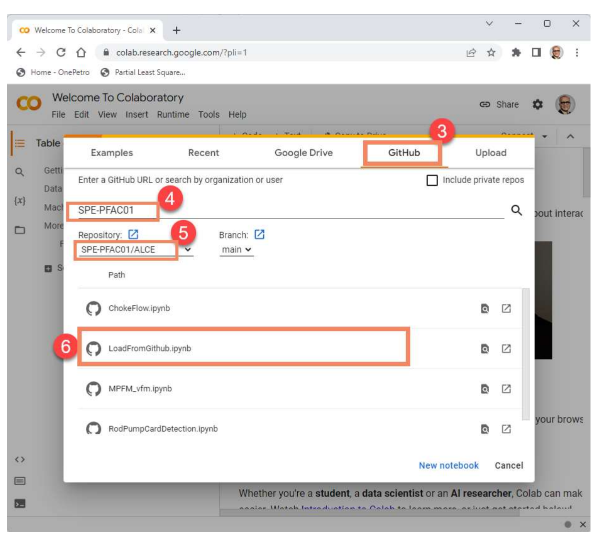
1. Once logged-in, open a browser (Chrome preferred but will work with Edge)
   – Go to https://colab.research.google.com/
2. Sign in into your Google account by clicking on the right top button.
   - You may setup another account just for the class if you don't want to use



5

Download first script from the Github

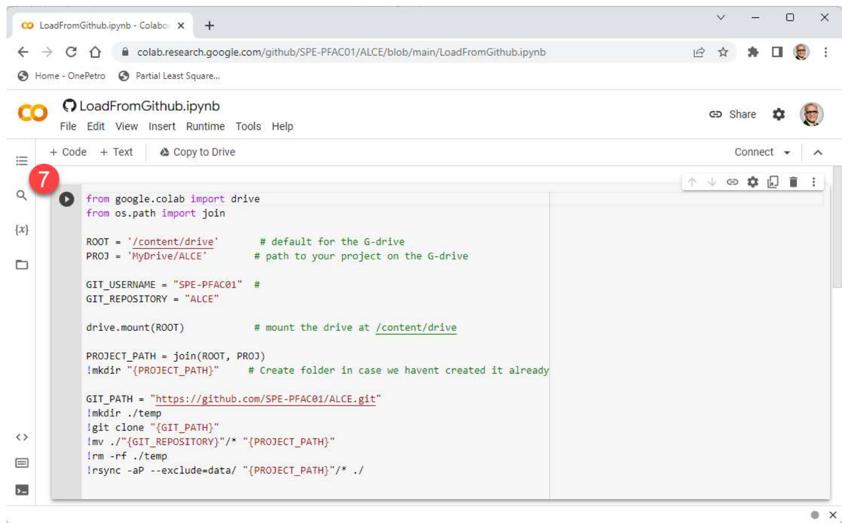
3. Click on 'GitHub' tab.
4. Type in 'SPE-PFAC01'
5. Select 'ALCE'
6. Double click on 'LoadFromGithub.ipynb' to download it into Colab.



6

Run the script to download other Python scripts & datastore

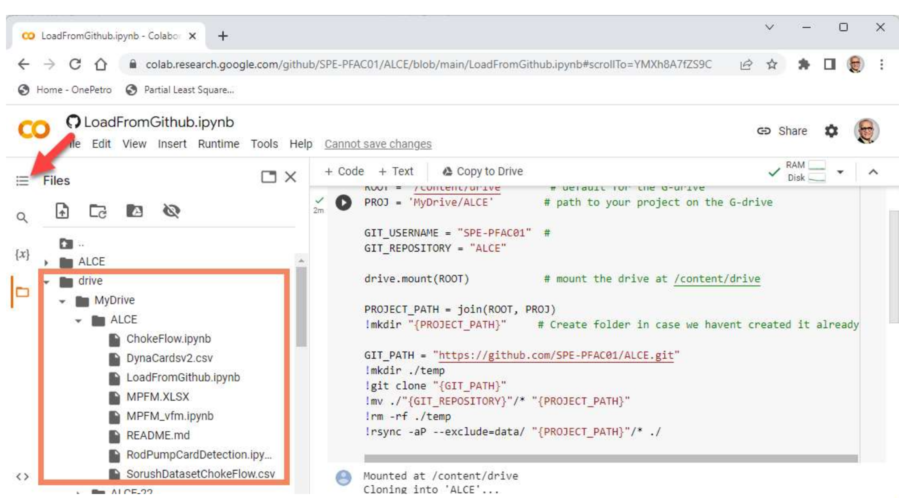
7. Click on the Run button.
8. On the Warning pop-up, click on 'Run Anyway'
9. On the next pop-up click on 'Connect to Google Drive'
10. Authenticate in another popup window by selecting 'Allow'.



7

Verify that you got it all!

- Click on the 'Files' symbol.
- Expand 'drive/MyDrive/ALCE' folder and check names of the eight files.



8

Pan American ENERGY

3.0. Diagnose Rod Pump Problems with AI/ML

Dr. Patrick Bangert
Algorithmica

Dr. Rajan Chokshi
Accutant Solutions LLC

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

SPE Western Regional Meeting | 23–26 April 2019 San Jose, California, USA | 2

Paper No. SPE-195295 Diagnosing and Predicting Problems with Rod Pumps using Machine Learning

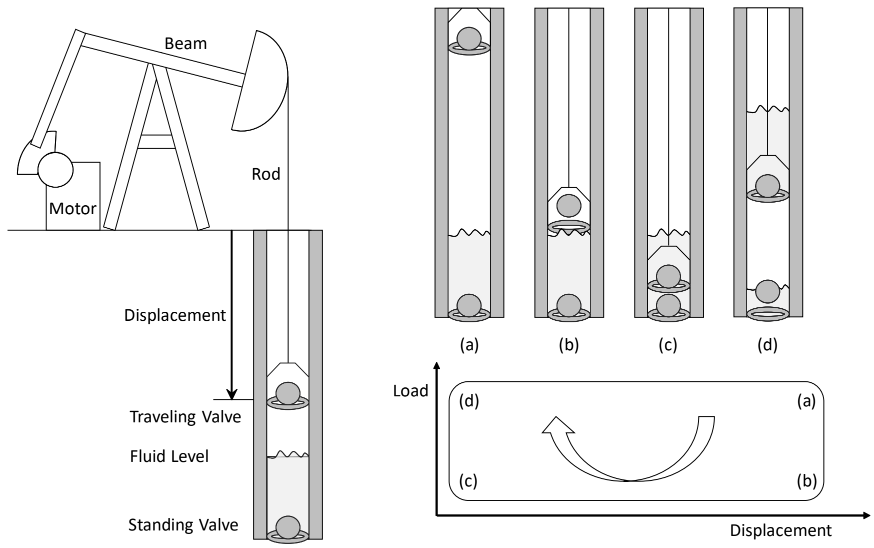
Patrick Bangert, algorithmica technologies
Sayed Sharaf, Tatweer Petroleum

algorithmica technologies
تطوير للبترول
TATWEER PETROLEUM

2

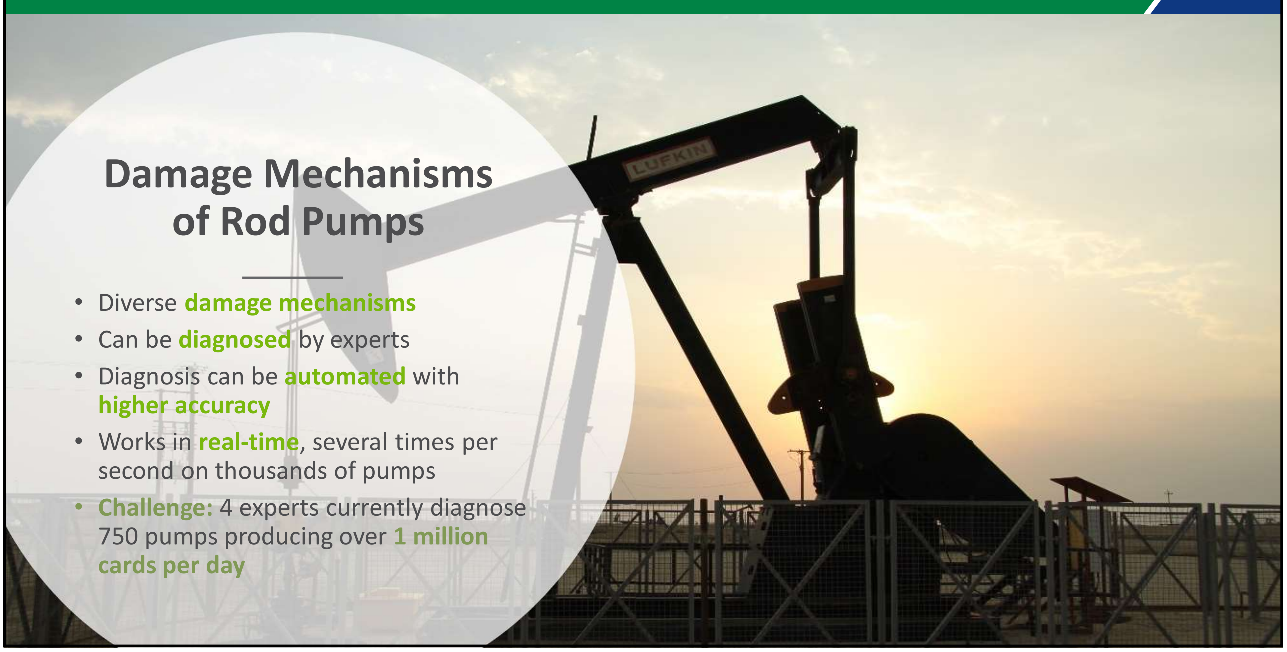
SPE Western Regional Meeting | 23–26 April 2019 San Jose, California, USA | 3

Rod Pump



3

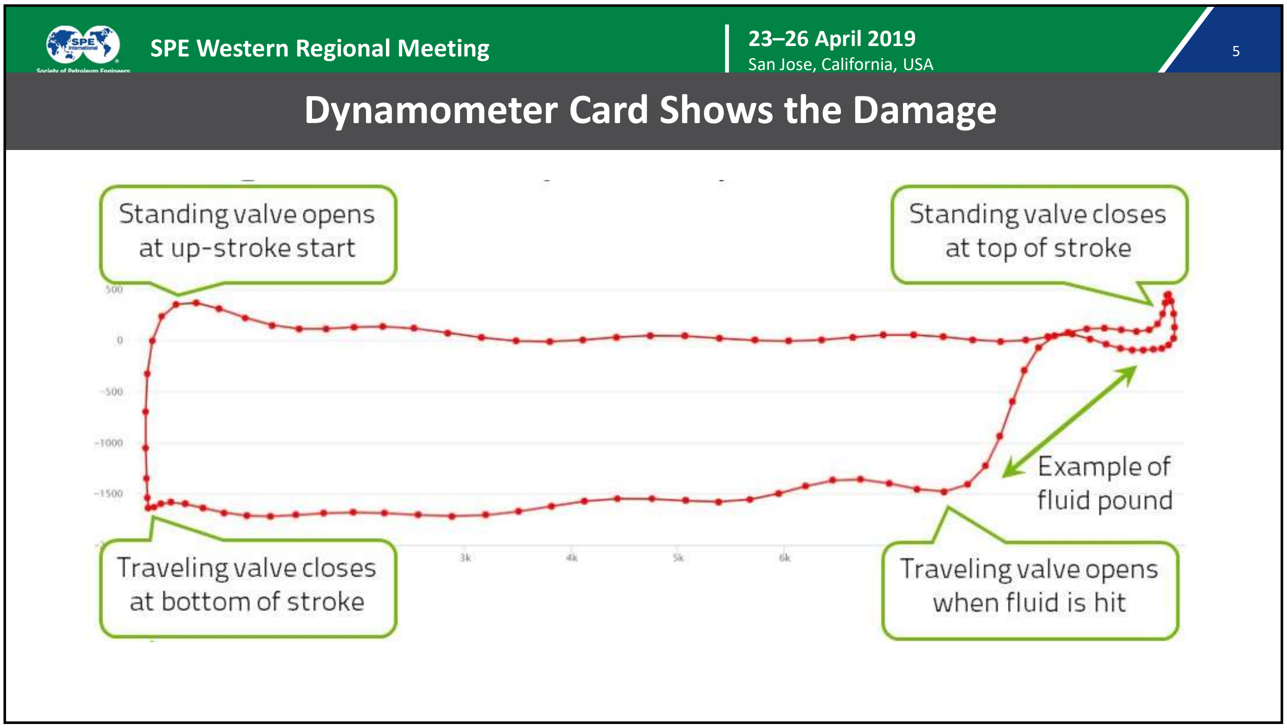
SPE Western Regional Meeting | 23–26 April 2019 San Jose, California, USA | 4



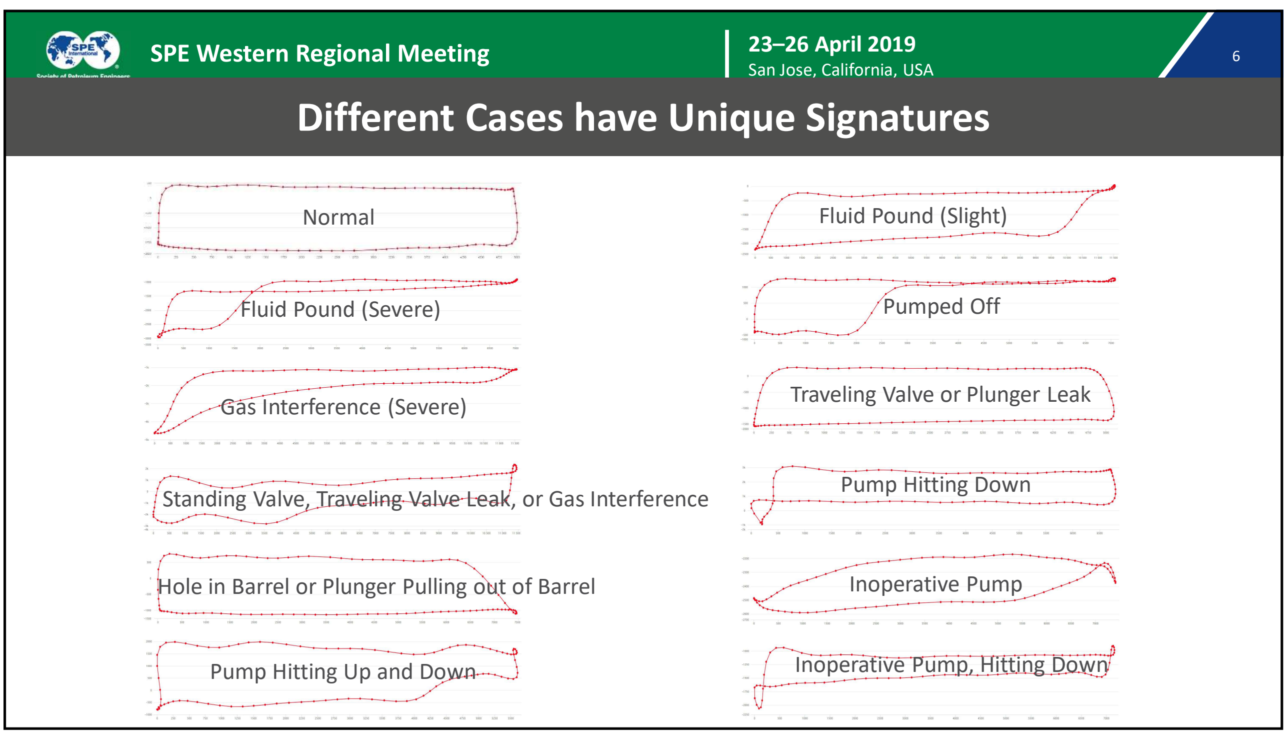
Damage Mechanisms of Rod Pumps

- Diverse **damage mechanisms**
- Can be **diagnosed** by experts
- Diagnosis can be **automated** with **higher accuracy**
- Works in **real-time**, several times per second on thousands of pumps
- **Challenge:** 4 experts currently diagnose 750 pumps producing over **1 million cards per day**

4


SPE Western Regional Meeting
23–26 April 2019
San Jose, California, USA
5
Dynamometer Card Shows the Damage
Standing valve opens at up-stroke start
Standing valve closes at top of stroke
Example of fluid pound
Traveling valve closes at bottom of stroke
Traveling valve opens when fluid is hit

5


SPE Western Regional Meeting
23–26 April 2019
San Jose, California, USA
6
Different Cases have Unique Signatures
Normal
Fluid Pound (Slight)
Fluid Pound (Severe)
Pumped Off
Gas Interference (Severe)
Traveling Valve or Plunger Leak
Standing Valve, Traveling Valve Leak, or Gas Interference
Pump Hitting Down
Hole in Barrel or Plunger Pulling out of Barrel
Inoperative Pump
Pump Hitting Up and Down
Inoperative Pump, Hitting Down

6

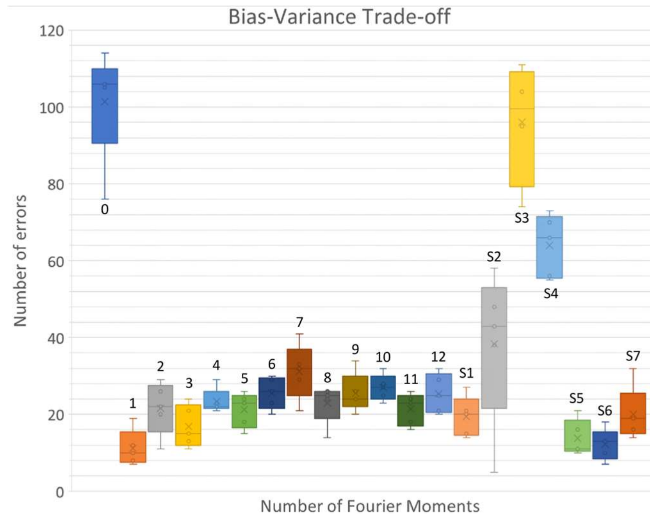
SPE Western Regional Meeting

23–26 April 2019
San Jose, California, USA

7

Feature Engineering

- Cards consist of 100 load and displacement measurements
- Need to have less variables
- Choose features that have best trade-off between variance and bias
  - Fourier series to one moment and centroid



7

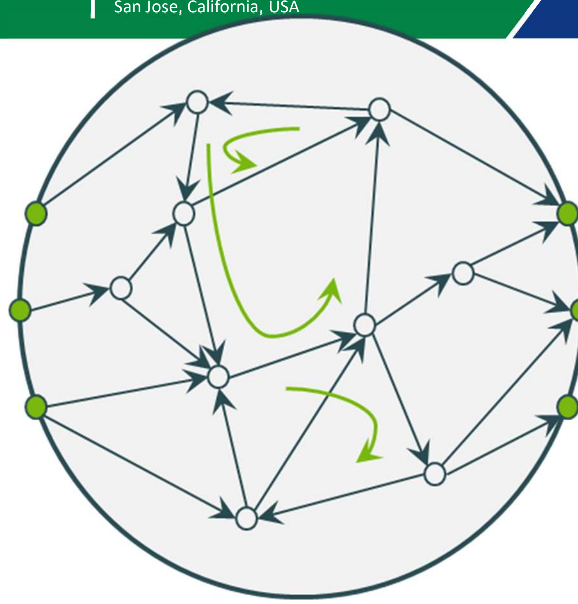
SPE Western Regional Meeting

23–26 April 2019
San Jose, California, USA

8

Damage Mechanisms of Rod Pumps

- Dynamometer Card is recorded **digitally**
- Known cards are turned into a **diagnosis model**
- Diagnosis takes place in the computer – user gets **email**
- **Maintenance** can be done immediately



8

SPE Western Regional Meeting | 23–26 April 2019
San Jose, California, USA | 9

Results

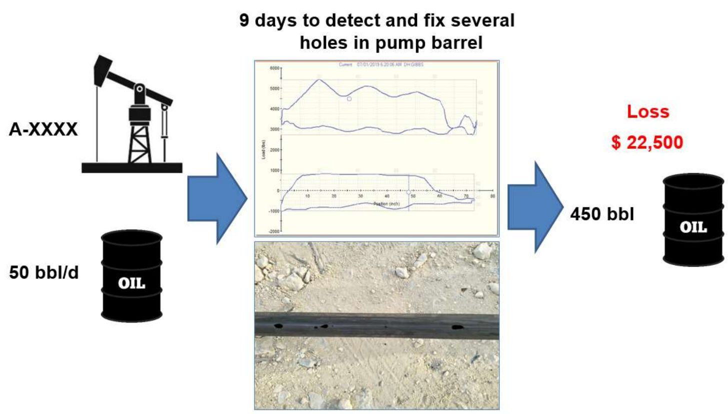
Category	Training Samples	Testing Samples	Incorrect
Normal	8557	1529	0
Fluid Pound (Slight)	5347	955	0
Fluid Pound (Severe)	93	15	0
Inoperative Pump	1981	379	0
Pump Hitting Down	1740	303	2
Pump Hitting Up and Down	2258	407	2
Inoperative Pump, Hitting Down	9045	1626	1
Traveling Valve or Plunger Leak	98	15	1
Standing Valve, Traveling Valve Leak, or Gas Interference	345	62	0
Pumped Off	234	39	1
Hole in Barrel or Plunger Pulling out of Barrel	132	20	0
Gas Interference (Severe)	101	11	0
Total	29931	5361	7

9

SPE Western Regional Meeting | 23–26 April 2019
San Jose, California, USA | 10

Added Value

- Some problems can be discovered (only) this way
- Problems are **discovered earlier** than before
- Team **focusses on maintenance** and not diagnosis
- Oil **production increases**

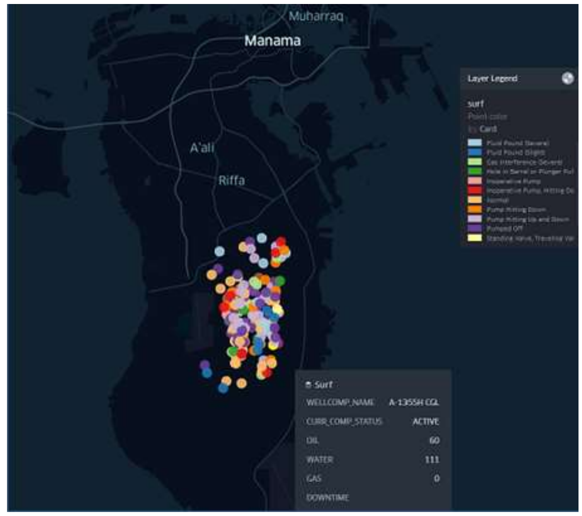


10

SPE Western Regional Meeting | 23–26 April 2019, San Jose, California, USA | 11

Next Steps

- Production is **controlled by parameters** that are set by experts
- Such as the **pump rate**
- Automatically **tune** these parameters
- **Optimize production** and maintenance together



11

Hands On Project

- Work with a similar dataset of dynamometer cards
  - 3370 Cards
  - Cards represent following conditions as labeled by SMEs
    - Flumping
    - Incomplete Fillage
    - Full Pump
    - Pump Hitting Down
- After data exploration, and feature analysis, we will develop and test several multi-class classification models with the following methods:
  - Decision Tree, Random Forrest, Support Vector Machine, Extra Trees, Gradient Boosting, XGBoost, and Neural Network.

12

Pan American ENERGY

4.0. Critical Choke Flow Rate Determination: Comparison Case Study

Dr. Rajan Chokshi

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

Outline

- Introduction
- Data Set Description
- ML Methods Evaluated
- Results
- Conclusion
- Hands-on Exercise

2

Introduction

Presentation is based on a publication:

- Barjouei, H.S., Ghorbani, H., Mohamadian, N. et al. **Prediction performance advantages of deep machine learning algorithms for two-phase flow rates through wellhead chokes**. J Petrol Explor Prod Technol 11, 1233–1261 (2021). https://link.springer.com/article/10.1007/s13202-021-01087-4 [Accessed 3 Jun 2021]
    - Note: The case study discusses data and findings from a SW Iranian oilfield published above.
    - This reference is selected because of detailed discussion on the methodologies, workflows, recent publication-timeline, and most-importantly the authors have made the entire dataset available.
    - All figures, tables in this presentation are from the above reference unless noted otherwise.
    - No rights are claimed.

3

Flow Rate Through Choke Problem

Control Volumes in Gas-Lift

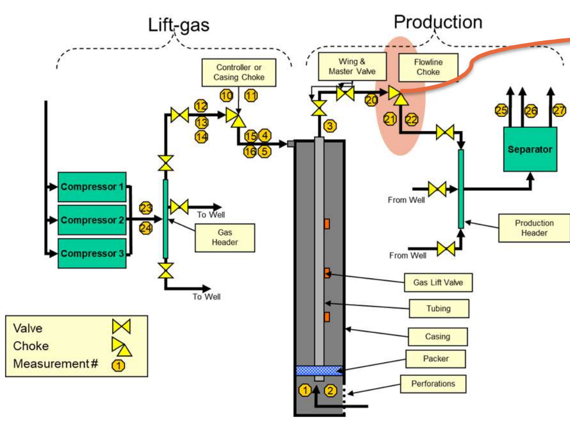


Image Source: POSC 2006

Production Choke

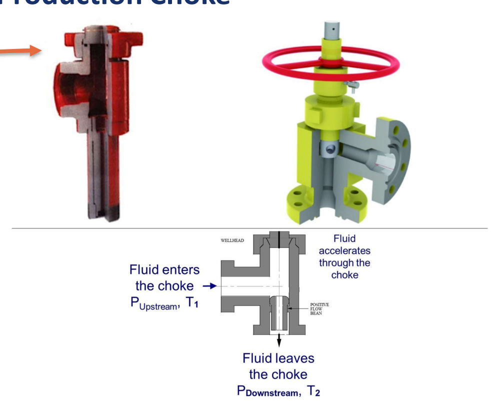
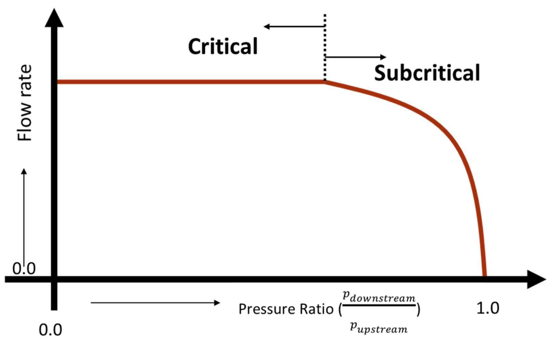


Image Source: Vendor literature

4

Flow Rate Through Choke

Flow Regimes in Choke



- The Critical flow occurs when the maximum fluid velocity inside the choke equals or exceeds the velocity of sound in the flowing fluid at in situ conditions.
- The flow rate will not change (remain constant) with further decrease in the downstream pressure, and it will be dependent only on the upstream conditions.

5

Empirical Expressions for Critical Flow Regime

Table 1 Empirical equations proposed for fluid flow across oil field chokes

Authors/year	Equation	Formula	Coefficient
Gilbert (1949)	(2)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c}{GLR^d}$	a = 0.1, b = 1, c = 1.89, d = 0.546
Baxendell	(3)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c}{GLR^d}$	a = 0.1046, b = 1, c = 1.93, d = 0.546
Ros	(4)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c}{GLR^d}$	a = 0.05747, b = 1, c = 2.00, d = 0.500
Achong	(5)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c}{GLR^d}$	a = 0.26178, b = 1, c = 1.88, d = 0.650
Pilehvari	(6)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c}{GLR^d}$	a = 0.021427, b = 1, c = 2.11, d = 0.313
Safar Beiranvand et al. (2012)	(7)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c (1-BS\&W\%)^d}{GLR^e}$	a = 0.0382, b = 1, c = 2.151, d = 0.52965, e = 0.5154
Ghorbani et al. (2019)	(8)	$Q_L = \frac{P_{wh}{}^b D_{64}{}^c (1-BS\&W\%)^d}{aGLR^e}$	a = 1.3522, b = 1, c = 1.7056, d = −0.164, e = 0.74042
Mirzaei-Paiaman et al. (case-4) (2013)	(9)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c \gamma_g^d \gamma_o^e}{GLR^f}$	a = 0.052439, b = 1, c = 1.9108, d = 0.3988, e = 0.1711, f = 0.5220
Choubineh et al. (2017)	(10)	$Q_L = a\frac{P_{wh}{}^b D_{64}{}^c \gamma_g^d \gamma_o^e \left(\frac{T}{T_{sc}}\right)^f}{GLR^g}$	a = 0:067,662, b = 1, c = 2:08,918, d = 0:625,862, e = 1:583,074, f = 0:000,453, g = 0:508,714

6

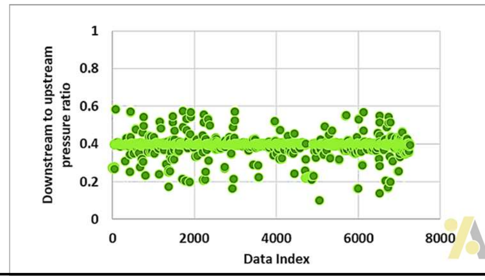
Dataset

Table 11 Data record statistical characterization of the variables in this study

Dataset Variables for Prediction Two-phase Flow Rate from Wellheads in the oil Fields

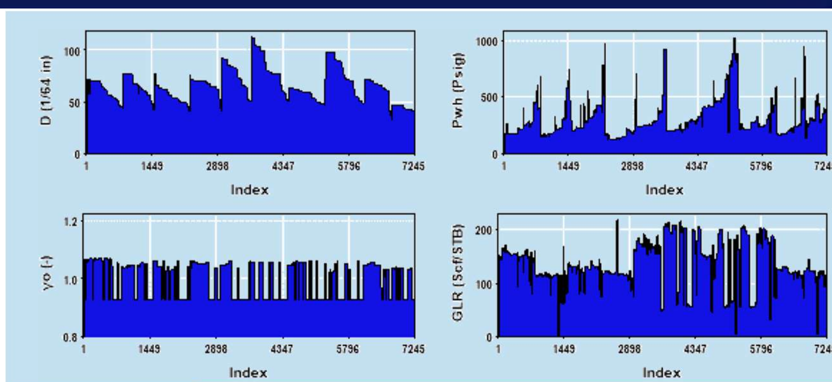
Field	Variables	Wellhead Choke Diameter	Wellhead Pressure	Oil Specific Gravity	Gas to Liquid Ratio	Two-Phase Flow Rate
	Symbol	D64	Pwh	Yo	GLR	QL
	Units	inch/64	psig	%	Scf/STB	STB/day
7245 dataset records from Sorush Oil Field	Mean	65.7	319	1.00	135.3	11,667
	Std. Deviation	15.2	169	0.06	42.9	3625
	Variance	231.5	28,718	0.00	1840.1	13,137,230
	Minimum	33.8	131	0.93	3.0	660
	Maximum	111.9	1024	1.07	217.0	23,700
113 dataset records from Oil Field of South Iran	Mean	54.5	1280	0.86	858.6	8549
	Std. Deviation	16.9	349	0.02	440.3	5376
	Variance	282.4	120,507	0.00	192,106.3	28,641,000
	Minimum	24.0	50	0.81	107.0	1324
	Maximum	80.0	2940	0.92	3660.0	22,150

- 10 wells from Sorush offshore field located in SW Iran
  - 7358 records with choke diameter, wellhead pressure, GLR, flow rate.
- Entire dataset in spreadsheet format is available here: **https://doi.org/10.1007/s13202-021-01087-4**



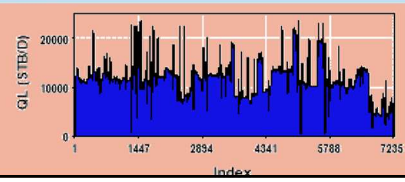
7

Data Visualization



Input Variables

Fig. 13 Variable value versus dataset index number highlighting the ranges and extreme values associated with each variable recorded for the Sorush oil field dataset. The variables displayed are: choke size (D64), wellhead pressure (Pwh), oil-specific gravity (yo), gas to liquid ratio (GLR), and two-phase flow rate through wellhead choke (QL) across all 7245 data records in Sorush Oil field

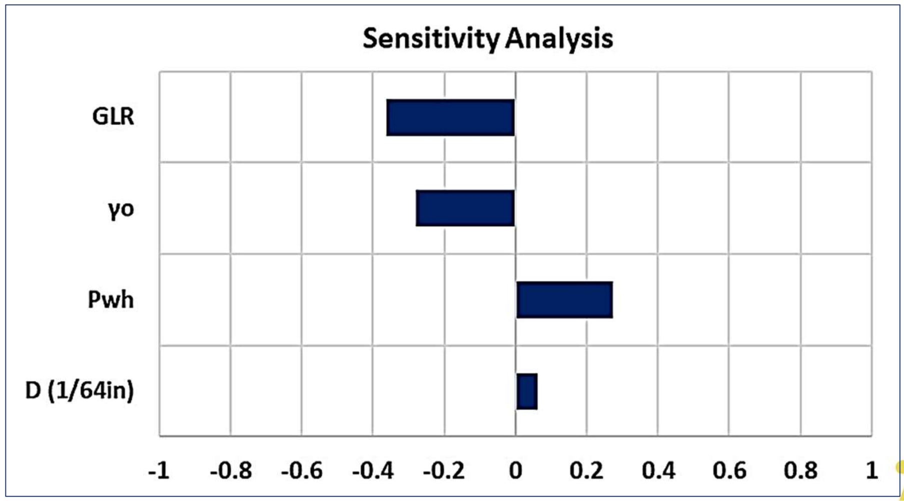


Dependent Variable

8

Data Exploration

- Spearman's correlation coefficients shown here reveal that the input variables,
  - GLR and γo are inversely related to QL,
  - Pwh and D64 display positive correlations with QL.
- D64 shows the lowest correlation coefficient with QL of the four input variables evaluated.
- Prevailing flow through the wellhead chokes of the Sorush oil field conforms to a critical flow regime



9

Workflow

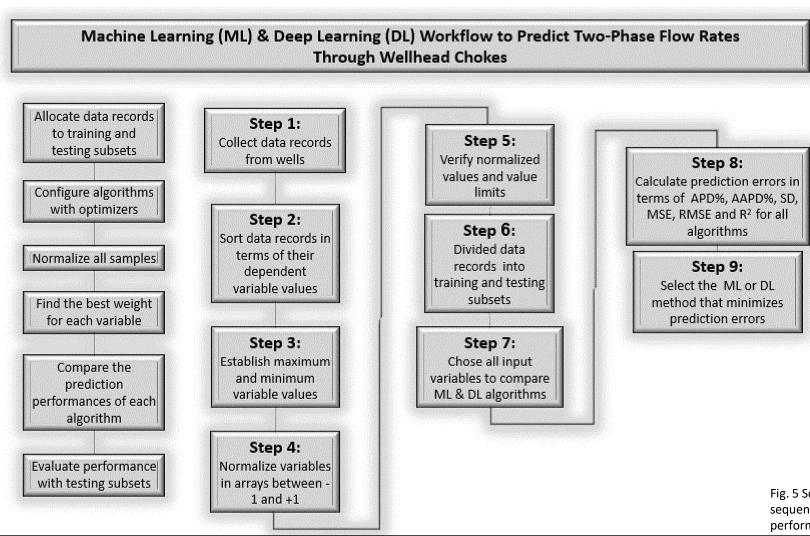


Fig. 5 Schematic diagram of the workflow sequence applied for comparing the prediction performance of ML and DL algorithms

10

Dataset Preparation

Normalization

begins with data collecting followed by data variable characterization, including the determination of the maximum and minimum values for each data variable involved. This information is used to normalize all data variables to range between + 1 and -1 by applying Eq. (11).

$$x_i^l = \left(\frac{x_i^l - xmin^l}{xmax^l - xmin^l} \right) * 2 - 1 \qquad (11)$$

where $x_i^l =$ the value of attribute for data record I;

$xmin^l =$ the minimum value of the attribute among all the data records in the dataset; and,

$xmax^l =$ the maximum value of the attribute among all the data records in the dataset.

Training & Testing Split

- 80% of the records to Training Dataset
- 20% of the records to the testing subset.

11

Learning Network Algorithms Selected

1. Support Vector Regression (SVR)
2. Decision Tree
3. Random Forest
4. Artificial Neural Network (ANN)
5. Deep Neural Network (DNN)

12

ML Model Control Parameters

1. Support Vector Regression

- Radial basis Function (RBF) Kernel used

Kernel function	Mathematical expression	Definition of parameters
Polynomial	$K(x,x_i) = \left(t + \frac{x_i^T x}{c}\right)^d$	d = degree of polynomial t = intercept
Sigmoid	$K(x,x_i) = \tanh(k x_i^T x + \theta)$	k= scale parameter θ = bias parameter
Radial basis function (RBF)	$K(x,x_i) = \exp\left(-\frac{\lVert x_i - x_j \rVert^2}{2\sigma^2}\right)$	σ^2= variance of RBF (Gaussian) kernel
Linear	$K(x,x_i) = x_i^T x$	

- Control Parameters

Parameters	Status
Kernel function	RBF
ε range	0.1
C range	100,000
Cross-validation	Not applied
γ range (RBF)	0.05

2. Decision Tree

- The scikit lean (Sklearn) Python module used.
- Control Parameters

Table 5 Decision tree wellhead choke flow regression model control parameter values

Parameters	Value
Maximum depth	100
Criterion	gini
Splitter	best
Objective function	Mean squared error (MSE)
Example prediction time	0.012925 (s)

13

ML Model Control Parameters

3. Random Forest

- Multiple decision trees in parallel, each using relatively few layers/nodes.

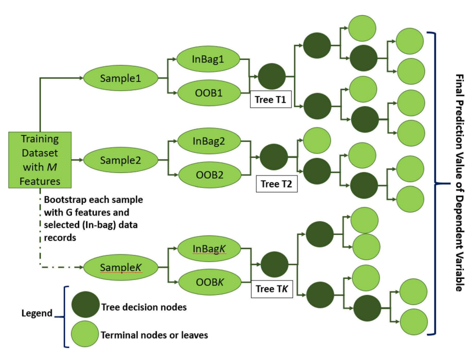


Table 6 Random forest wellhead choke flow regression model control parameter values

Parameters	Value
Maximum depth	1000
Random state	0
Number of decision trees	1000
Objective function	Mean squared error (MSE)
Example prediction time	7.352307 (s)

14

ML Model Control Parameters

4. Artificial Neural Network (ANN)

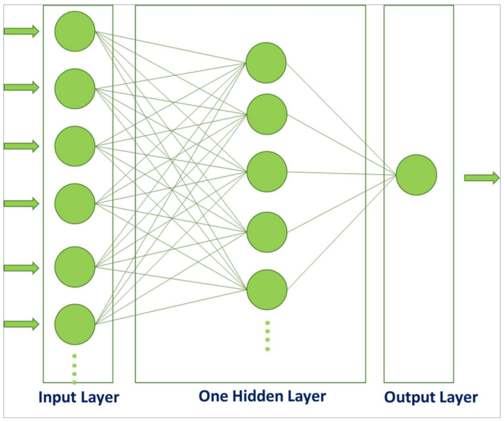


Table 7 Control parameters for the one-hidden layer ANN model constructed to predict two-phase flow rate (Q_l) through wellhead choke

Control Parameters	Status
Number of hidden layers	1
Number of neurons in the hidden layer	500
Activation function used input to hidden layer	SELU (Scaled Exponential Linear Unit)
Activation function used hidden to output layer	SELU
Objective function minimized for training subset	MSE
Optimization algorithm	RMSprop
Minimum delta	0.01% of Q_L mean value
Patience (number of iterations)	25
Number of iterations	237
Learning rate	0.01

15

ML Model Control Parameters

The Most Common Activation Function

Table 8 The most common activation functions

Activation function	Formula
Sigmoid	$s(x) = \frac{1}{1+e^{-x}}$
tanh	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Relu (Rectified linear unit)	$y(x) = \max(0, x)$
Selu (Scaled exponential linear unit)	$y(x) = \lambda \begin{cases} x & if x > 0 \\ \alpha e^x - \alpha & if x \leq 0 \end{cases}$
Leaky-relu	$y(x) = \max(ax, x) : 0 < a \ll 1$
Linear	$y(x) = x$

Common Objective Functions

Objective function	Abbreviation	Formula
Mean of squared error	MSE	$J(y, \hat{y}) = \frac{1}{n}\sum_i^n (\hat{y}_i - y_i)^2$
Mean of absolute error	MAE	$J(y, \hat{y}) = \frac{1}{n}\sum_i^n (\lvert\hat{y}_i - y_i\rvert)$
Mean of absolute error%	MAPE	$J(y, \hat{y}) = \frac{1}{n}\sum_i^n \frac{\lvert\hat{y}_i - y_i\rvert}{y_i} * 100$
Mean of squared logarithmic error	MSLF	$J(y, \hat{y}) = \frac{1}{n}\sum_i^n (\log(\hat{y}_i + 1) - \log(y_i + 1))^2$

16

ML Model Control Parameters

5. Deep learning Neural Network (ANN)

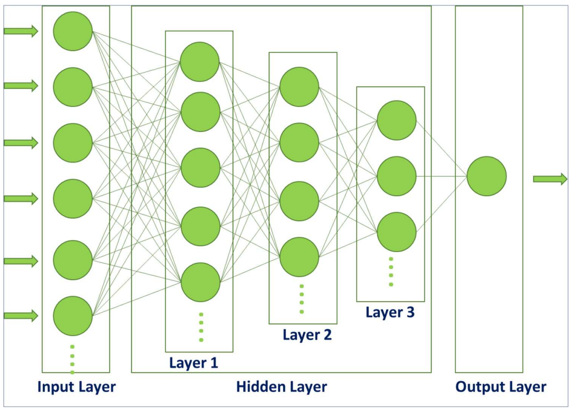


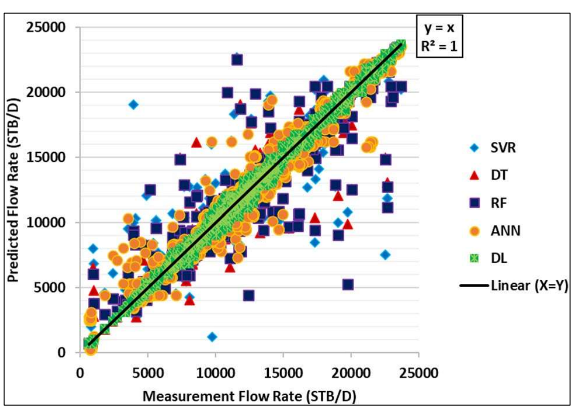
Table 10 Control parameters for the multi-layer DL model constructed to predict two-phase flow rate (Q_L) through wellhead choke

Control Parameters	Status
Number of hidden layers	3
Number of neurons in hidden layers, 1, 2, and 3	500, 100, 50
Activation function used in input / hidden layers	SELU
Activation function applied to output layer	SELU
Objective function used training	MSE
Optimization algorithm	RMSprop
Minimum delta	0.01% of Q_L mean value
Patience (number of iterations)	25
Number of iterations	418
Learning rate	0.1

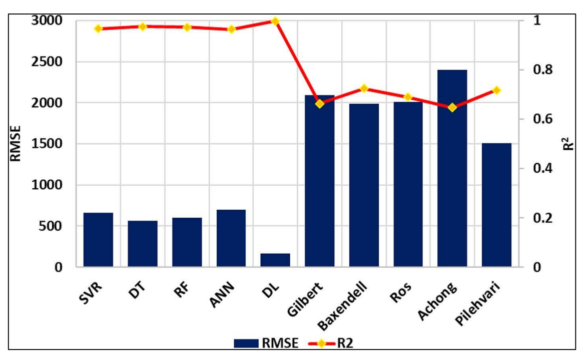
17

Results

Predicted vs Measured Flow Rates



Flow Rate Prediction Accuracy



18

Results

Two-phase liquid flow rate prediction accuracy for the total subset (7265 Data records; 100%)

Models		APD	AAPD	SD	MSE	RMSE	R^2
Units		(%)	(%)	(STBD)	(STBD)	(STBD)	
SVR	Support vector regression	−0.570	1.780	659.3	434,679	659.3	0.9670
DT	Decision tree	−0.245	1.780	566.6	321,106	566.7	0.9756
RF	Random forest	−0.306	1.665	600.1	360,155	600.1	0.9726
ANN	Artificial neural network	−0.038	5.396	697.9	487,019	697.9	0.9644
DL	Deep learning	−0.019	0.757	168.8	28,481	168.8	0.9978
Math-ematical Models	Gilbert	3.620	11.482	2196.0	4,370,300	2090.5	0.6631
	Baxendell	2.120	11.024	1259.3	3,959,439	1989.8	0.7244
	Ros	2.561	11.781	1681.5	4,052,142	2013.0	0.6894
	Achong	0.449	12.419	1579.0	5,740,491	2395.9	0.6467
	Pilehvari	−0.211	10.392	1230.6	2,280,496	1510.1	0.7178

Table 14 QL Prediction performance compared for ML, DL, and traditional mathematical models applied to the entire dataset (7245 data records from Sorush oil field) for the Sorush field dataset of wellhead choke recordings

Percentage difference (PD):

$$PD_i = \frac{\xi_{(Measured)} - \xi_{(Predicted)}}{\xi_{(Measured)}} x100 \qquad (24)$$

Average percent deviation (APD):

$$APD = \frac{\sum_{i=1}^{n} PD_i}{n} \qquad (25)$$

Absolute average percent deviation (AAPD):

$$AAPD = \frac{\sum_{i=1}^{n} |PD_i|}{n} \qquad (26)$$

Standard deviation (SD):

$$SD = \sqrt{\frac{\sum_{i=1}^{n} (PD_i - Dimean)^2}{n-1}} \qquad (27)$$

$$Dimean = \frac{1}{n} \sum_{i=1}^{n} (\xi_{Measured\ i} - \xi_{Predicted\ i}) \qquad (28)$$

Mean square error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\xi_{Measured\ i} - \xi_{Predicted\ i})^2 \qquad (29)$$

Root-mean-square error (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} \qquad (30)$$

where n = number of data records; x_i = measured dependent variable value for the i^{th} data record; and y_i = predicted dependent variable value for the i^{th} data record.

Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\xi_{Predicted\ i} - \xi_{Measured\ i})^2}{\sum_{i=1}^{N} \left(\varphi_{Predicted\ i} - \frac{\sum_{i=1}^{n} \xi_{Measured\ i}}{n}\right)^2} \qquad (31)$$

19

Conclusions

- Comprehensive study on the critical flow regime in oil wells
- Sizeable dataset was used
- Multiple ML and NN methods applied and evaluated using several criteria.
  - Deep-learning Neural Network performed the best followed by Decision Tree, Random Forest and Support Vector Regression.
  - Simple Neural network performed behind traditional ML methods.
  - All empirical methods perform poorly.

20

Pan American ENERGY

5.0. Data-driven Flow Pattern Prediction In A Nearly Horizontal Pipeline: A Comparison

Presented by: Dr. Rajan Chokshi
Case & Script prepared by: Gerardo Vera
Advisor: Dr. Eduardo Pereyra, TUHWALP

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

1

Outline

- Introduction
- Data Set Description
- ML Methods Evaluated
- Results
- Conclusion

2

Introduction

- Main aim:
  - Reduce time required to determine flow patterns in a relatively accurate manner.
- It's a classification problem
  - In Classification, a program learns from the dataset or observations and then classifies new observation into a number of classes or groups.
- Reference:
  - Pereyra et al. "A methodology and database to quantify the confidence level of methods for gas–liquid two-phase flow pattern prediction," Chemical Engineering Research and Design, 2012, https://doi.org/10.1016/j.cherd.2011.08.009.

3

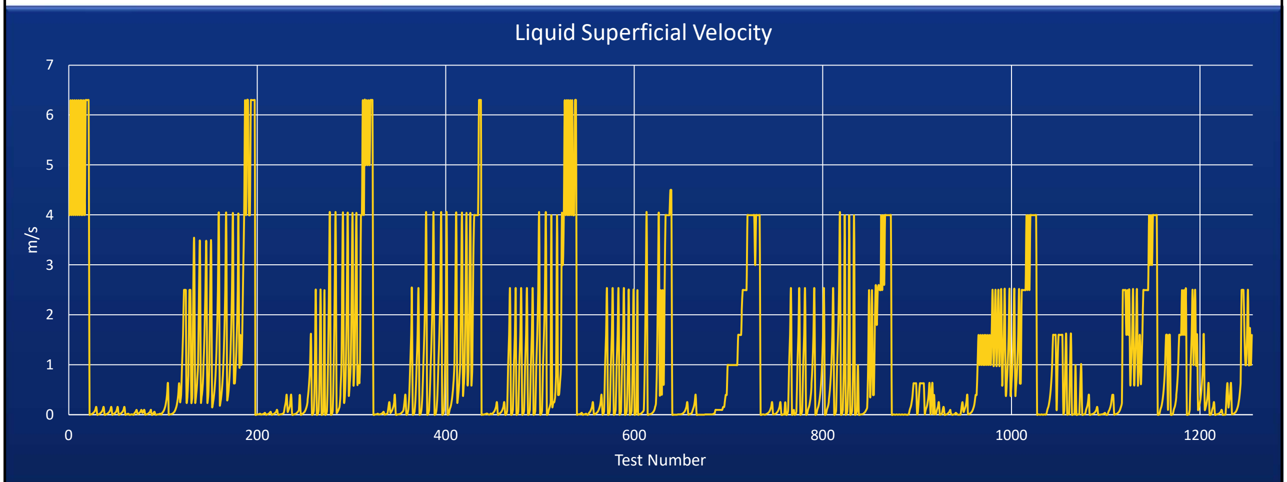
Dataset

- The earliest set of data is Shoham (1982), which was acquired in 50.8 and 25.4 mm pipe diameters, utilizing air water at atmospheric conditions. This was the first study covering systematically all the inclinations angles, from −90◦ to +90◦.

	Input Data														Output	
	Test Number		Material Properties							System Geometry			Operational Cond		Unnamed: 17_level_1	Unnamed: 18_level_1
	Test Code	P	Type of liquid	Type of Gas	DenL	DenG	VisL	VisG	ST	ID	Roughness	Ang	Vsl	Vsg	Flow Pattern	Flow Pattern
0	1982_Ovadia Shoham_LD_Water_Aira1	101353.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	4.0	0.025	DB	1
1	1982_Ovadia Shoham_LD_Water_Aira2	101354.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	6.3	0.025	DB	1
2	1982_Ovadia Shoham_LD_Water_Aira3	101355.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	4.0	0.040	DB	1
3	1982_Ovadia Shoham_LD_Water_Aira4	101356.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	6.3	0.040	DB	1
4	1982_Ovadia Shoham_LD_Water_Aira5	101357.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	4.0	0.063	DB	1
5	1982_Ovadia Shoham_LD_Water_Aira6	101358.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	6.3	0.063	DB	1
6	1982_Ovadia Shoham_LD_Water_Aira7	101359.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	4.0	0.100	DB	1
7	1982_Ovadia Shoham_LD_Water_Aira8	101360.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	6.3	0.100	DB	1
8	1982_Ovadia Shoham_LD_Water_Aira9	101361.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	4.0	0.160	DB	1
9	1982_Ovadia Shoham_LD_Water_Aira10	101362.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	6.3	0.160	DB	1
10	1982_Ovadia Shoham_LD_Water_Aira11	101363.268601	water	Air	1000.0	1.8	0.001	0.000015	0.07	0.051	0	0.0	4.0	0.400	DB	1

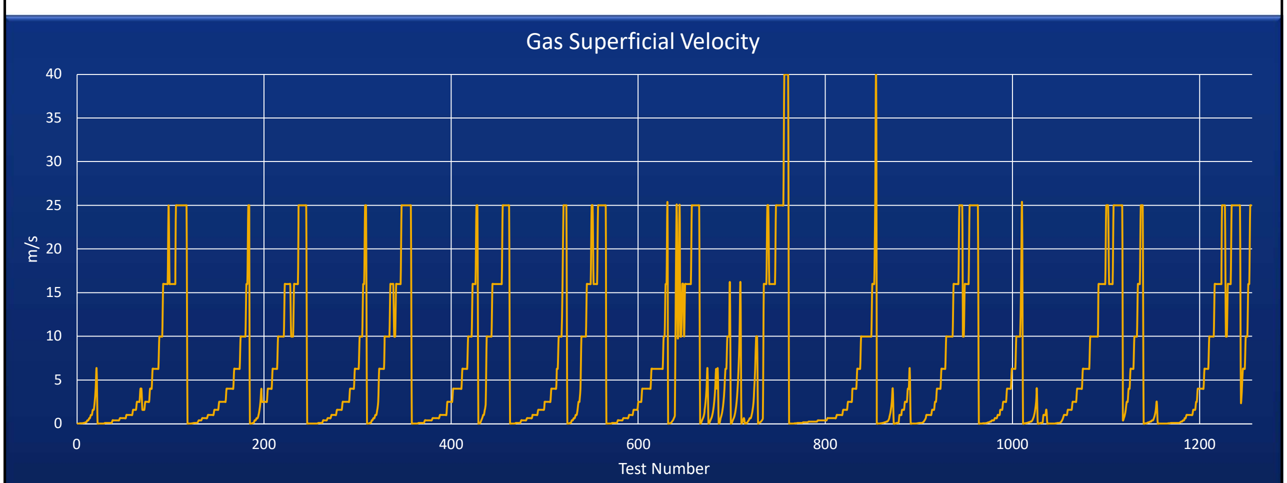
4

Input data



5

Input data



6

Input data

7

Constant parameters

- Liquid and gas density
- Liquid and Gas viscosity
- Superficial tension
- Pipe Inner diameter (2 inch)
- Pipe roughness

8

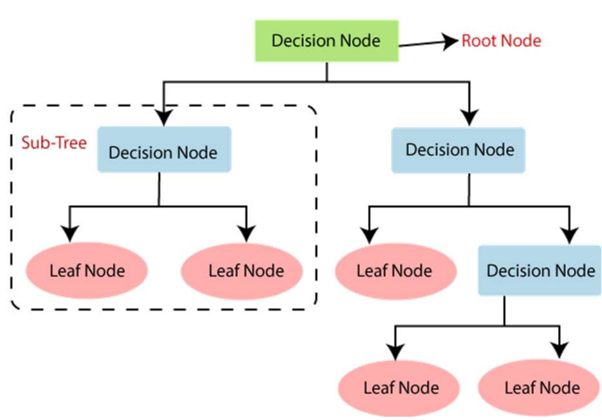
Machine learning techniques evaluated

- Decision Tree Classifier
- Random Forest Classifier
- Naïve Bayes classifier
- Support Vector Machine [SVM]

9

Decision Tree Classifier

- The simplest yet the most versatile ML technique
  - A graphical representation for getting all the possible solutions to a problem/decision based on given conditions.
  - Our problem's decision tree ended up 10-levels deep with 76 leaves.
- Supervised learning
- Classification, binary
  - reduce entropy (randomness amount)
- Risk of overfitting



10

Random Forest Classifier

- Random forest algorithm
  - Supervised learning
  - Classification
  - Multiple decision trees reduces overfitting, overcomes missing data

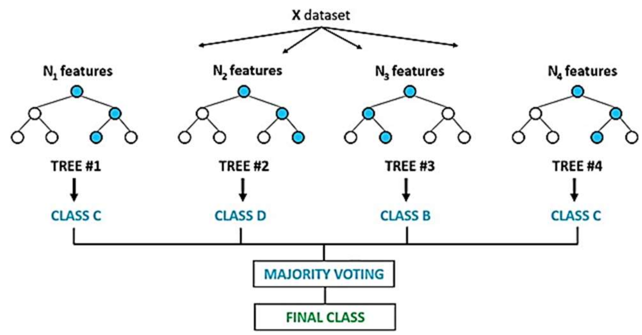


Image: https://www.youtube.com/watch?v=goPiwckWE9M

11

Naive Bayes Classifier

- Naive bayes classifier
  - Supervised learning
  - A Probability Classifier using Bayes theorem.
    - Called 'Naïve' because it requires a strong assumption of independence between input variables.
  - Goal is to calculate Conditional probability for each of possible outcomes (in our case five outcomes are possible)

$$P(A|B) = \frac{P(B|A)\, P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$

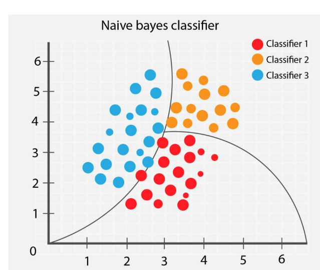
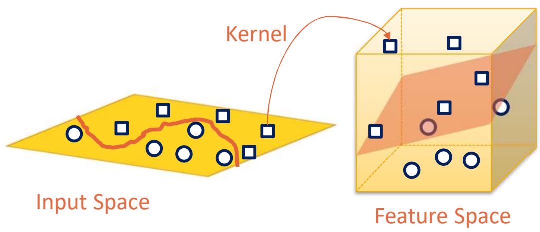


Image Source: https://towardsdatascience.com/introduction-to-naive-bayes-classifier-fa59e3e24aaf

12

Support Vector Machine

- Support vector machine
  - Supervised learning
  - Classification, less common regression, geometrical split, kernel function to increase dimensions



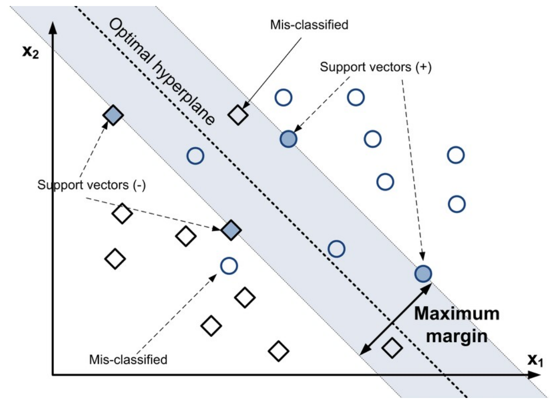


Image Source: https://www.researchgate.net/figure/Concept-of-Support-Vector-Machines-The-optimal-hyperplane-separates-two-classes-of_fig1_315784930

13

Python Code for Fitting the models

```
#Decision Tree Algorithm
modelDTA = DecisionTreeClassifier(criterion = "entropy", random_state = 100)
modelDTA.fit(data2SVM, data3)
#Random Forest Algorithm
modelRFA = RandomForestClassifier(n_jobs = 2, random_state = 0)
modelRFA.fit(data2SVM, data3)
#Naive Bayes Classifier
modelNBC = GaussianNB()
modelNBC.fit(data2SVM, data3)
#Support Vector Machine
modelSVM = svm.SVC(C = 100)
modelSVM.fit(data2SVM, data3)
```

Parameter Explanatory:

- **Entropy** : criterion of split in this case information gain, alternative gini for the impurity
- **Random_state** : The features are always randomly permuted at each split to choose the one to split at certain node
- **N_jobs** : Number of jobs to run in parallel
- **Random_state** : As above, not important better to be fixed to reproduce results
- **GaussianNB** : main formula to be applied, in this case this is the best (numerical data), others are for categorical data. Others are, Gaussian, Multinomial (text), Complement, Bernoulli, Categorical.
- **C** : Regularization parameter, the bigger the most accurate but the most time consuming as well.

14

Predicting the results

Python Code

```python
#Decision Tree Algorithm
predictDTA = modelDTA.predict(data5SVM)
print(predictDTA[0:25])
print("")
#Random Forest Algorithm
predictRFA = modelRFA.predict(data5SVM)
print(predictRFA[0:25])
print("")
#Naive Bayes Classifier
predictNBC = modelNBC.predict(data5SVM)
print(predictNBC[0:25])
print("")
#Support Vector Machine
predictSVM = modelSVM.predict(data5SVM)
print(predictSVM[0:25])
```

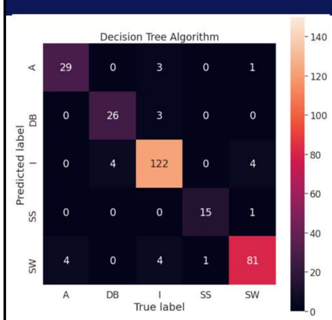
Predictions

```
['DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'SS' 'SS' 'SS' 'SS'
 'SS' 'SS' 'SS' 'SS' 'SS' 'SW' 'SW' 'SW' 'SW' 'SW' 'SW']

['DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'SS' 'SS' 'SS' 'SS'
 'SS' 'SS' 'SS' 'SS' 'SS' 'SW' 'SW' 'SW' 'SW' 'SW' 'SW']

['DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'SS' 'SS' 'SS' 'I' 'I'
 'SS' 'I' 'SS' 'SS' 'SS' 'SW' 'SW' 'A' 'A' 'A']

['DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'DB' 'SS' 'SS' 'SS' 'SS'
 'SS' 'SS' 'SS' 'SS' 'SS' 'SW' 'SW' 'SW' 'SW' 'SW' 'SW']
```

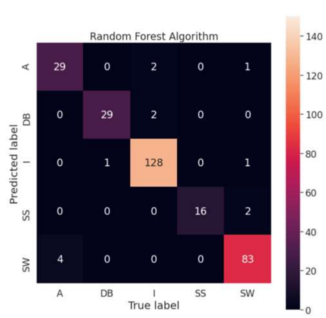
A – Annular **DB – Dispersed Bubble** **I – Intermittent** **SS – Stratified Smooth** **SW – Stratified Wavy**

15

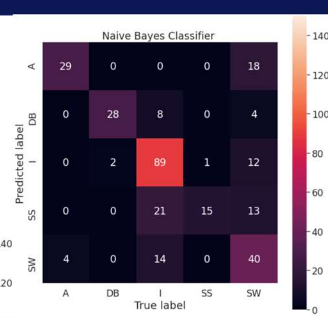
Results – Confusion Matrix for Each ML Method



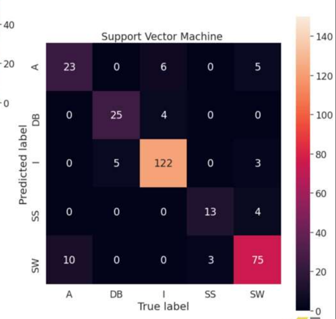
Decision Tree



Random Forest



Naive Bayes Classifier



Support Vector Machine

16

Comparing methods

```python
title = ['Decision Tree Algorithm Resulst:', 'Random Forest Algorithm Results:',
         'Naive Bayes Classifier Results:', 'Support Vector Machine Results:']
accuracy = [predictDTA, predictRFA, predictNBC, predictSVM]
for x, y in zip(title, accuracy):
    str = '\033[1m' + x + '\033[0m'
    print('')
    print(str)
    print('')
    print('Accuracy:', "%.2f" % (accuracy_score(data6, y)* 100))
```

Decision Tree Algorithm Resulst:

Accuracy: 91.61

Random Forest Algorithm Results:

Accuracy: 95.64

Naive Bayes Classifier Results:

Accuracy: 67.45

Support Vector Machine Results:

Accuracy: 86.58

Algorithm	Accuracy, %
Decision Tree	91.61
Random Forest	95.64
Naïve Bayes Classifier	67.45
Support Vector Machine	86.58

17

Conclusions

- This study shows that the best machine learning method *for this purpose, among the evaluated ones*, is the **Random Forest** algorithm. The Decision Tree algorithm is right behind.
- This study confirms what is known from experience and theory, that the flow pattern can be accurately predicted if the superficial velocities are known, and the inclination angles are considered.
- This study may be even more useful if done using dimensionless parameters.

18

More detailed References on this topic

- Ezzatabadipour M et al. (2017). Deep learning as a tool to predict flow patterns in two-phase flow. https://arxiv.org/abs/1705.07117 [Accessed 3 Jun 2021]
- Guilen-Rondon et al. (2018). Support Vector Machine Application for Multiphase Flow Pattern Prediction. https://arxiv.org/pdf/1806.05054.pdf [Accessed 3 Jun 2021]

19

Pan American ENERGY

6.0. Well Test Analysis

Dr. Rajan Chokshi

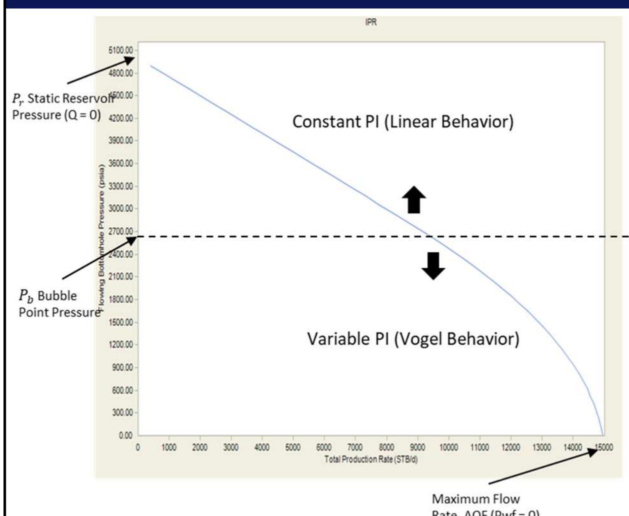
Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

Productivity Determination



- Well's productivity, Inflow Performance Relationship – IPR, is helpful
  - Operating a well safely while honoring the capacity of
    - Reservoir
    - Surface equipment
    - Artificial lift equipment
  - Designing for conditions as reservoir and surface operations change
    - Completion changes
    - Artificial lift changes
    - Stimulation
    - EOR
- Well tests provide snapshot on well's productivity.

2

Well Test Types

Test Type	Measurement Condition*		Characteristics
	Flowing	Shut-In	
Production Well Test	✔	✔	Stabilized test in an established production or injection well
Production Log Test	✔	✔	Downhole measurement of pressure, flow rate, temperature and (usually) density
Multi-rate Test	✔		Measurements for a succession of surface flow rates
Flowing Gradient Survey	✔		Downhole measurements of pressure, temperature with surface flow measurements.
Static Gradient Survey		✔	Downhole measurement of pressure, temperature and usually density.
Liquid Level Test	✔	✔	Downhole pressure measured or computed from liquid level soundings

*The well-tests are performed in stabilized or pseudo-steady-state flowing or shut-in conditions.

3

3

Production Well Test Report for a multi-rate test

Unit					
Pump Co.		Type	ESP	Model:	Opt.Range 825-2000

13 02

Flow period/ String	Start date / time yyyy.mm.dd hh:mm	End date / time yyyy.mm.dd hh:mm	Choke Size /64"	Casing Status	WHP psig	WHT deg F	GC Line Pressure psig	Critical flow Y / N	Sep. Temp. deg F	Sep. Pressure psig	Total Rate bbls/day	Oil Q bbls/day	Water Q bbls/day	Gas Q MMSCF/da
1	/02/13 08:15	0/02/13 09:15	48+00	Unit	169	129	138	N	46	154	1369	322	1047	0.044
2	/02/13 09:15	0/02/13 13:45	48+00	GC	173	130	140	N	69	158	1395	328	1067	0.051
3	/02/13 13:45	0/02/13 14:15	00+192	Unit	142	124	140	N	71	141	N/A	N/A	N/A	0.000

General comments	[illegible] has been flowing artificially by ESP through adjustable choke 00+26/64". Tubing well head pressure was 165 psig and Flow line was 140 psig.Started first flow period through fixed choke 00+48/64" (matching initial pressure) and modified 3" line (192/64") from casing to separator. In flow period # 2 Diverted casing to GC and continue flowing tubing through separator .Flow period # 3, diverted tubing flow to GC and casing flow divertod through Separator. Note :Observed no flow from casing .The test was done as per instructions from [illegible] Artificial lift Engineer [illegible] .The data given here are the average readings of the stable flow period. At 14:15 End of PGOR test and bypass separator.
Casing Existing Condition	Flowing to GC. In flow period # 1 casing flowed to the separator. At 10:15 am casing diverted to GC.In flow period # 3 Casing flowed to Separator

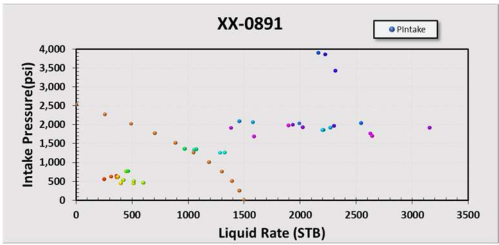
GOR 1	GOR 2	Total GOR	BS&W	SG OIL	OIL	SG GAS	Total GLR	Pi	Pd	Ti	Tm	Lift Type	ESP	
												Frequency	Voltage	Current
SCF/bbl	SCF/bbl	SCF/	%	@ 60 deg F	Deg API	(air = 1)	SCF/bbl	psia	psia	deg F	deg F	Hz	Volt	Am
136	49	185	76.5	0.892	27.2	0.846	32	1013	1864	134	198	50	1113	29
154	50	204	76.5	0.892	27.2	0.846	36	1010	1843	134	199	50	1113	29
N/A	N/A	N/A	N/A	N/A	N/A	0.846	N/A	1012	1828	134	199	50	1113	29

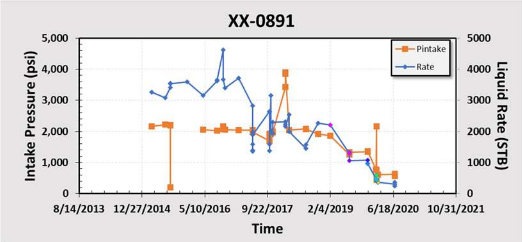
4

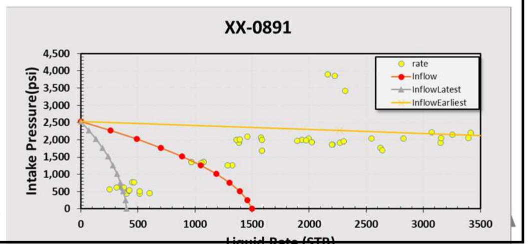
Challenge of Productivity Determination

date	rate	WHP	Pintake
6/30/2020 12:15	360	137	638
6/30/2020 10:15	252	137	560
6/30/2020 9:00	313	146	623
2/19/2020 12:00	369	165	603
2/19/2020 11:00	364	167	606
2/19/2020 10:00	370	169	605
2/19/2020 8:30	374	166	615
2/5/2020 18:00	399	255	449
2/5/2020 17:00	514	199	450
7/6/2019 10:15	1326	266	1264
7/6/2019 8:30	1286	256	1257
2/4/2019 10:15	2206	420	1860
2/4/2019 8:15	2200	424	1859
10/28/2018 10:00	2269.66	470.89	1915
3/11/2018 11:10	1992	418	2035
3/11/2018 9:30	2544	428	2036
2/9/2018 11:15	2163	400	3896
2/9/2018 8:30	2224	394	3857
2/8/2018 20:45	2315	117	3422
5/24/2017 13:15	2824	62	2040
2/2/2017 8:15	3711	326	2048
10/20/2016 8:15	3393	447	2051
10/1/2016 22:00	3664	415	2077
10/1/2016 19:00	4622	71	2161
8/11/2015 14:00	3540	462	2198
8/11/2015 13:00	3544	442	195
8/11/2015 11:15	3415	439	2205
7/1/2015 8:15	3075	496	2221
3/15/2015 8:00	3256	441	2156

- Well's productivity changes over time as reservoir is depleted and/or enhanced oil recovery processes are applied.

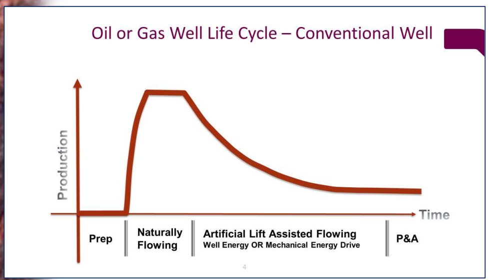


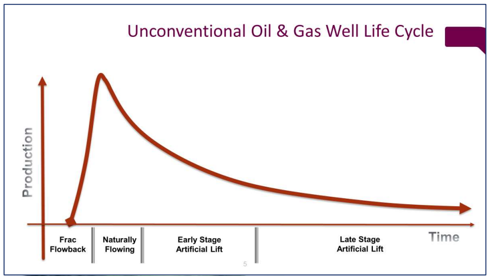


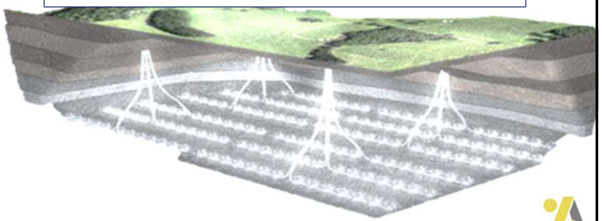


5

Unconventional Lifecycle is different





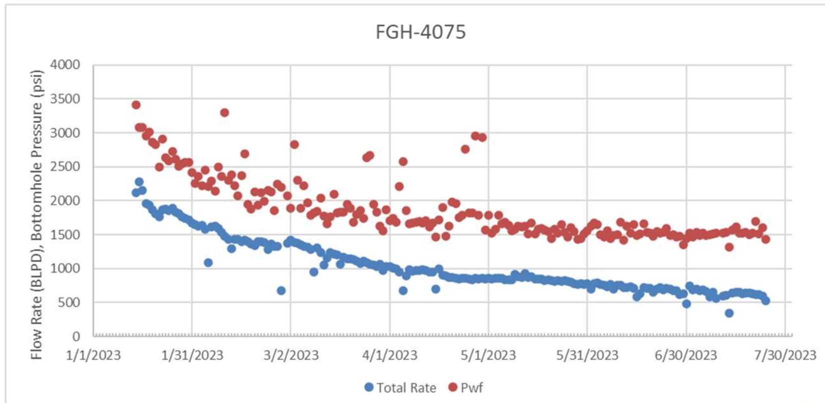


6

Challenge of Productivity Determination

- In unconventional, productivity changes are drastic, particularly in an earlier part of the well's life cycle.

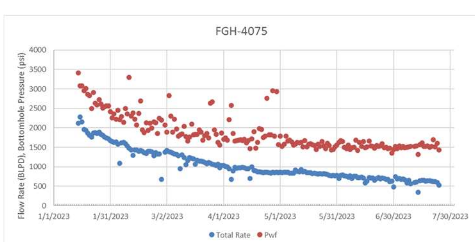
Test Date	Total Rate	Oil	Water	Gas	WC	Pwf
7/24/2023	522	244	278	408	53.3	1430.3
7/23/2023	598	288	310	425	51.8	1598.4
7/21/2023	619	291	328	87	53	1689.9
[illegible]	[illegible]	294	[illegible]	457	[illegible]	[illegible]
6/27/2023	676.2	321.2	355	492.5	52.5	1408.5
6/26/2023	674	311	363	442	53.9	1499.1
6/24/2023	707.8	333.1	374.7	199.1	52.9	1592.1
6/23/2023	685	324	361	445	52.7	1535.8
[illegible]	[illegible]	338	[illegible]	[illegible]	[illegible]	[illegible]
6/3/2023	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
6/2/2023	785	369	416	511.6	53	1672.7
6/1/2023	704	324	380	288	54	1625.5
5/29/2023	781	375	406	470	52	1446.3
5/28/2023	766	365	401	476	52.3	1428.9
4/11/2023	980	478	502	678	51.2	1673.5
4/10/2023	976	475	501	676	51.3	1695
4/5/2023	677	316	361	359	53.3	2375.7
4/4/2023	954	456	498	593	52.2	2209.4
3/14/2023	1243	600	643	836	51.7	1759
3/13/2023	1158	554	604	728	52.2	1658
3/6/2023	1328	634	694	291	52.3	2222.1
3/5/2023	1353	647	706	876	52.2	1890.9
2/17/2023	1393.2	651.6	741.5	896.3	53.2	1943.2
2/16/2023	1424	663.3	760.6	900.4	53.4	2696.7
2/15/2023	1403	657.6	745.4	786.2	53.1	2370.9
2/8/2023	1593.8	733	860.8	975	54	2492.2
2/7/2023	1626.6	749	877.6	1015	54	2138.8
1/22/2023	[illegible]	861	[illegible]	1022	53.9	[illegible]
1/21/2023	1760.8	796.8	964	1019.5	54.7	2495.3
1/20/2023	1805.9	811.5	994.4	1023.1	55.1	2830.7
1/10/2023	[illegible]	940.3	[illegible]	[illegible]	55	[illegible]
1/15/2023	2277.7	1023.5	1254.2	1292.3	55.1	3080.8
1/14/2023	2113.5	935.5	1178	1135.7	55.7	3409.4

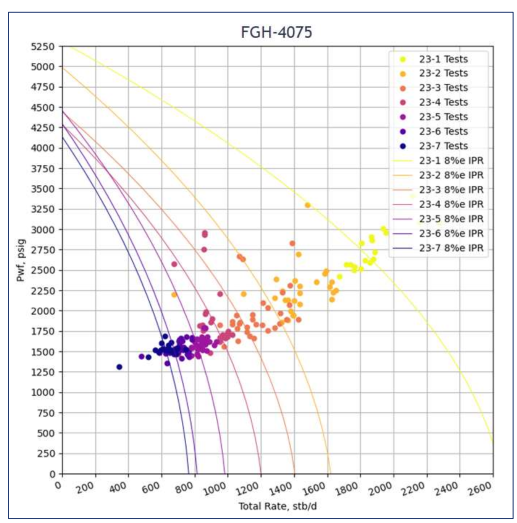


7

How do you determine IPR parameters?

- SBHP & PI both vary simultaneously.
- Considerable data volume





8

Workflow

- Import Well tests usually from a CSV or an excel file.
- Select a group of well test
  - Conventional well – Usually same month or quarter tests
  - Unconventional with high frequency
    - 3-days or 5-days
- Use scipy.optimize to minimize errors between calculated and measured flow rates for each group, assuming constant Pb, and varying SBHP, PI, and AOFP.
  - Calculate the flow rate for each data set using an unsaturated Vogel model.
- Select SBHP, PI values within error tolerance.

9

Workflow Demonstration

10

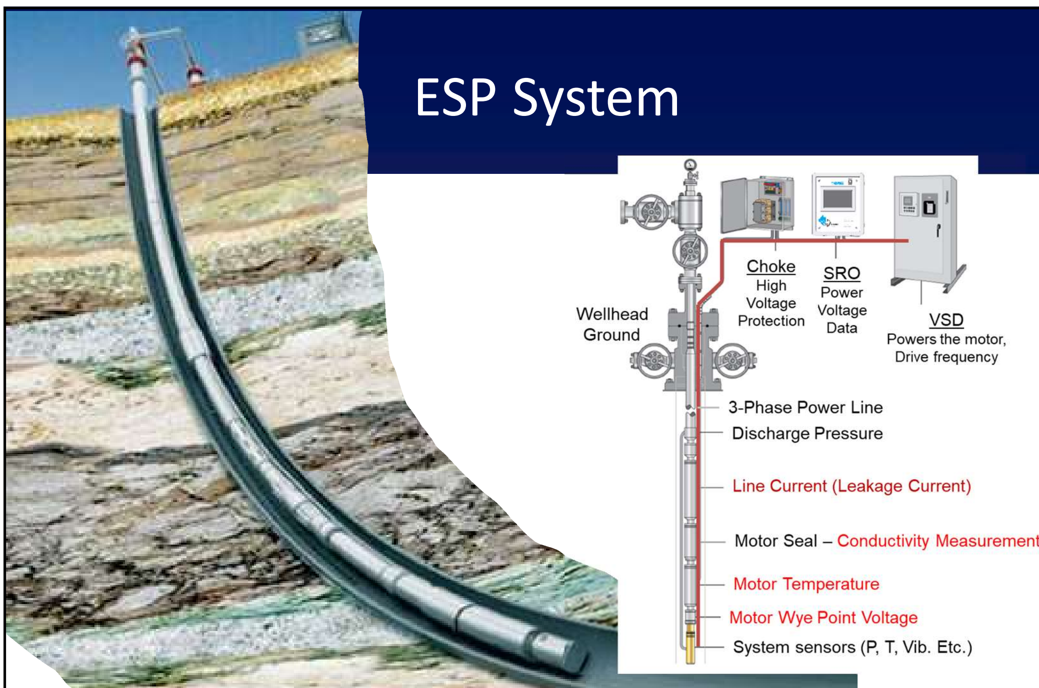
7.0. ESP Failure Analysis Review

Dr. Rajan Chokshi

Data Analytics for Artificial Lift & Production Optimization

1

ESP System



2

Critical Real-time Measurements

Surface Measurements
- Wellhead Pressure
- Casing Pressure
- Flowline Pressure
- Choke Opening
- Frequency

Downhole Measurements
- Intake Pressure
- Intake Temperature
- Discharge Sub Pressure
- Vibration (Vx, Vy, Vz)
- Line Voltage
- Motor Temperature
- Line Current (Leakage Current)
- Wye Point Voltage
- Motor Fluid Conductivity

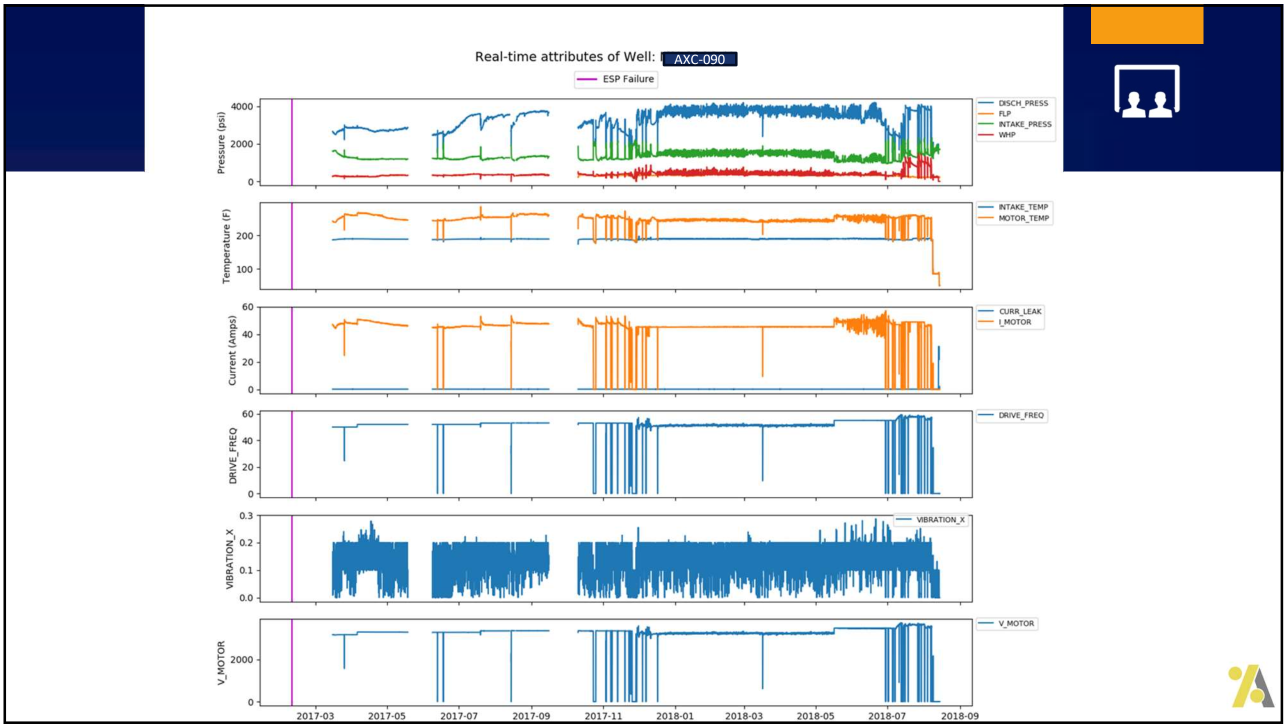
Measurements shown in **RED** are indicators of the Electrical Health of the ESP System.

3

Non-Real-time measurements are equally important

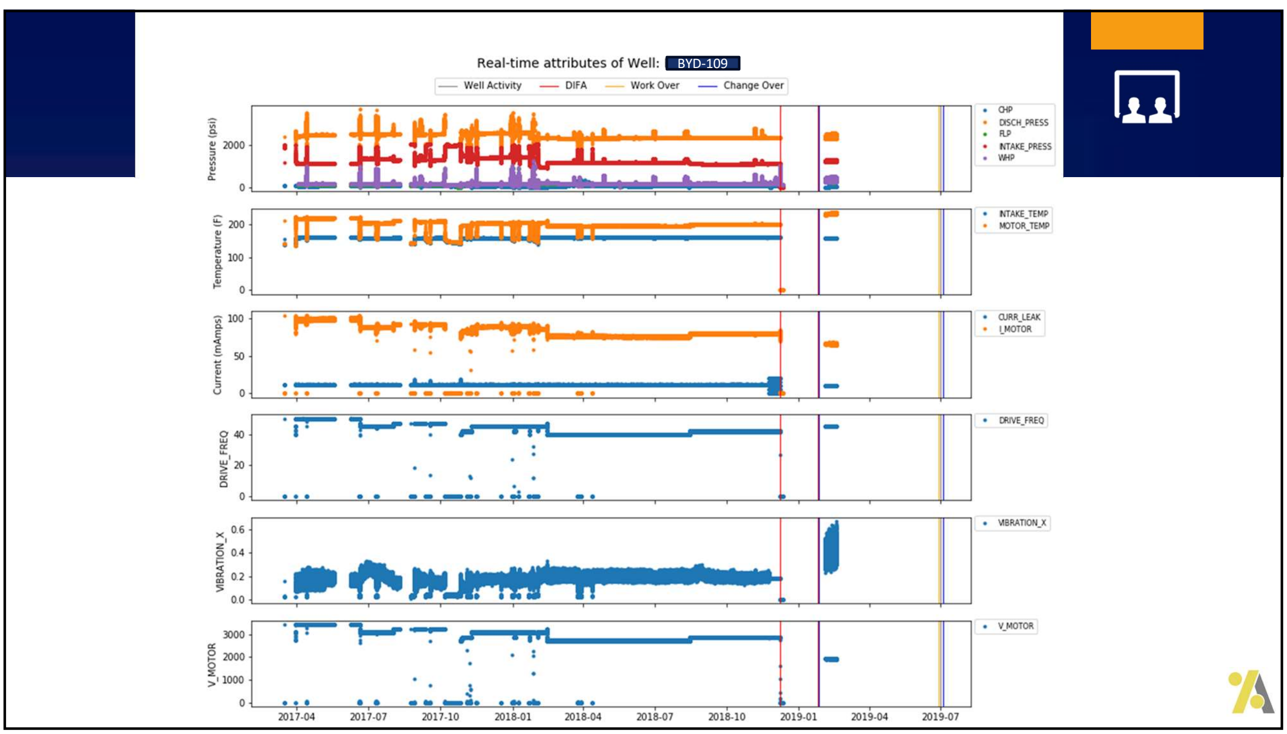
- Well tests
- ESP Equipment
  - History
    - Any component reused?
    - Run-time and time-from-installation
  - DIFA – Dismantle, Failure Analysis
- Well activity reports – including servicing work
- DIFA reports

4



Real-time attributes of Well: AXC-090
ESP Failure
Pressure (psi)
DISCH_PRESS
FLP
INTAKE_PRESS
WHP
Temperature (F)
INTAKE_TEMP
MOTOR_TEMP
Current (Amps)
CURR_LEAK
I_MOTOR
DRIVE_FREQ
VIBRATION_X
V_MOTOR
2017-03
2017-05
2017-07
2017-09
2017-11
2018-01
2018-03
2018-05
2018-07
2018-09

5



Real-time attributes of Well: BYD-109
Well Activity
DIFA
Work Over
Change Over
Pressure (psi)
CHP
DISCH_PRESS
FLP
INTAKE_PRESS
WHP
Temperature (F)
INTAKE_TEMP
MOTOR_TEMP
Current (mAmps)
CURR_LEAK
I_MOTOR
DRIVE_FREQ
VIBRATION_X
V_MOTOR
2017-04
2017-07
2017-10
2018-01
2018-04
2018-07
2018-10
2019-01
2019-04
2019-07

6

Failure Prediction for ESP using ML – Case Study 1

ESP Well and Component Failure Prediction in Advance using Engineered Analytics - A Breakthrough in Minimizing Unscheduled Subsurface Deferments
- Introduction
- Analytical Data Modeling
  - Machine Learning
  - Approach and Process
- Data Preparation
  - Data Gathering
  - Data Pre-Processing
- Model Building
  - Dataset Sample Selection
  - Model Training
  - Feature Selection and Engineering
  - Parameter Tuning
- Model Testing
  - Creating a Dataset for Model Testing
  - Model Scoring
  - Test Results
- Model Evaluation by Well Owners (Operator's Blind Test)
  - Methodology
  - Operator's Blind Test Results
- Value Generated by the Analytical Model
- Conclusions and the way ahead
- References

Static Data
- Installation and Commissioning data
- Well Completion information
- ESP Well Failure history
- Ad-hoc information (emails, spreadsheet, reports, photos)

Dynamic Data
- Downhole Gauge: Pump Discharge & Intake pressures, Motor temperature, Intake temperature, vibration etc.
- Motor Controller: Electrical readings, well running status and trip conditions.
- Surface Pressure Gauges: Tubing head pressure, Flow line pressure

Calculated Data
- Well Model Based Calculated Data: Inferred production and Pump Efficiencies calculated from the Physics based models in Real Time

Supervised learning approach: Random Forest based classifier developed using above and engineered features like standard deviations, param differences across fixed time intervals, differential pressures, etc.

Well Level Failure and component level failure models developed.

Source: Marin *et al.* (2019) ESP Well and Component Failure Prediction in Advance using Engineered Analytics - A Breakthrough in Minimizing Unscheduled Subsurface Deferments, SPE-197806

7

Failure Prediction for ESP using ML – Case Study 1

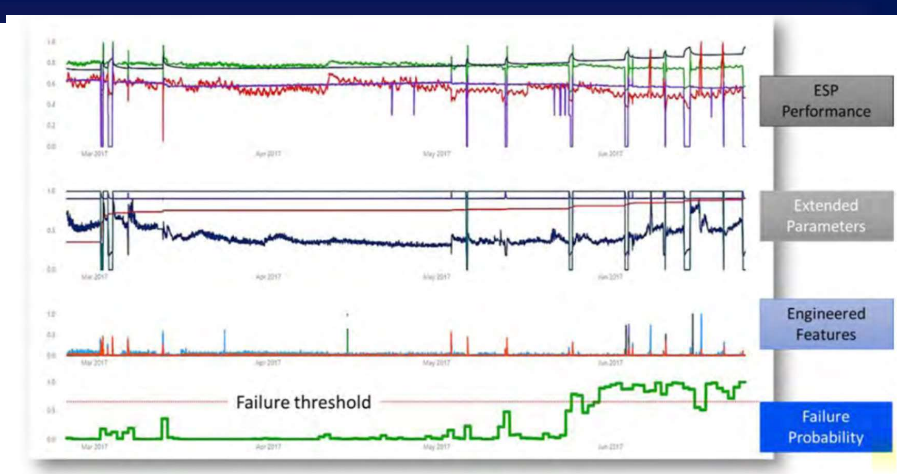


Figure 8—Captured Case (True Positive) where Well Failure Probability crosses the threshold and it is an actual Failure event

Source: Marin *et al.* (2019) ESP Well and Component Failure Prediction in Advance using Engineered Analytics - A Breakthrough in Minimizing Unscheduled Subsurface Deferments, SPE-197806

8

Failure Prediction for ESP using ML – Case Study 1

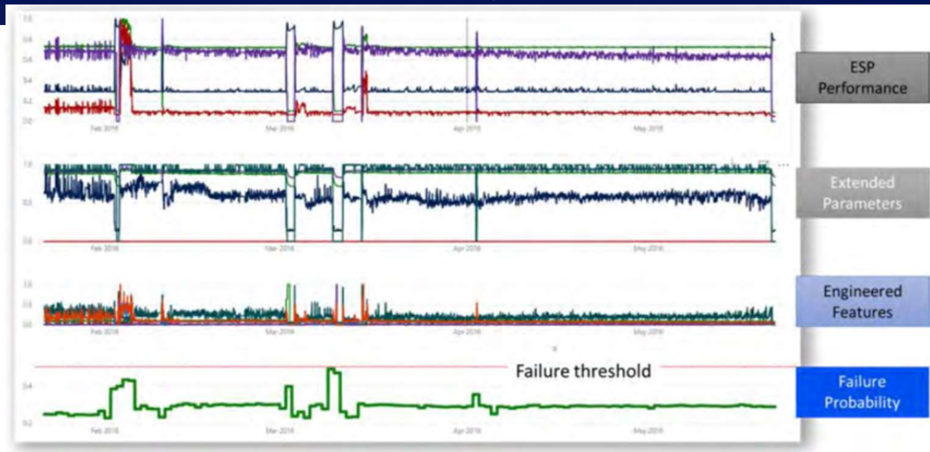


Figure 9—Missed Case (False Negative) where Well Failure Probability does not cross the threshold however it is an actual Failure event

Source: Marin *et al.* (2019) ESP Well and Component Failure Prediction in Advance using Engineered Analytics - A Breakthrough in Minimizing Unscheduled Subsurface Deferments, SPE-197806

9

Failure Prediction for ESP using ML – Case Study 2

Source: Silvia, Shejuti *et al.* (2022). "Case Study: Predicting Electrical Submersible Pump Failures Using Artificial Intelligence." Paper presented at the Offshore Technology Conference, Houston, Texas, USA, May 2022. doi: https://doi.org/10.4043/31852-MS

Table 1—Data collected to develop Predictive Failure Analytics (PFA)

	Source	Type of Data	Data
(a)	ESP Tracking Spreadsheet	ESP Operations and Manufacturing	ESP specifications, run number, installation date, start date, pull date, failure date, field name, well name, event log
(b)	DIFA Reports	ESP Reliability Data	ESP pull reason, failure description, root cause, failure component, comments from teardown
(c)	SCADA/ Historian	Surface and Downhole Sensor Data, Well Test Data	Drive frequency, motor current, voltage, intake pressure, discharge pressure, wellhead pressure intake temperature, motor temperature, current leakage, vibration, total fluid production rate, water-cut, GOR

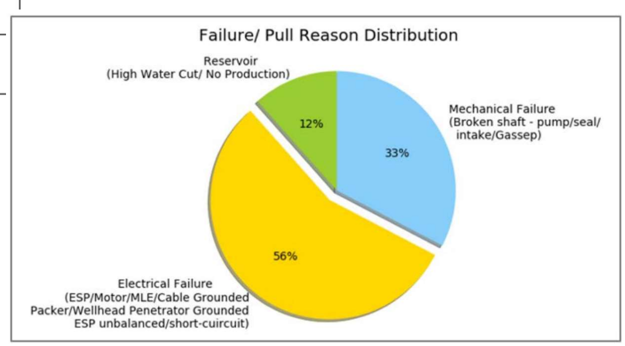


Figure 2—Failure/Pull reason distribution of 101 failed pumps from three producing assets in the Americas.

10

Failure Prediction for ESP using ML – Case Study 2

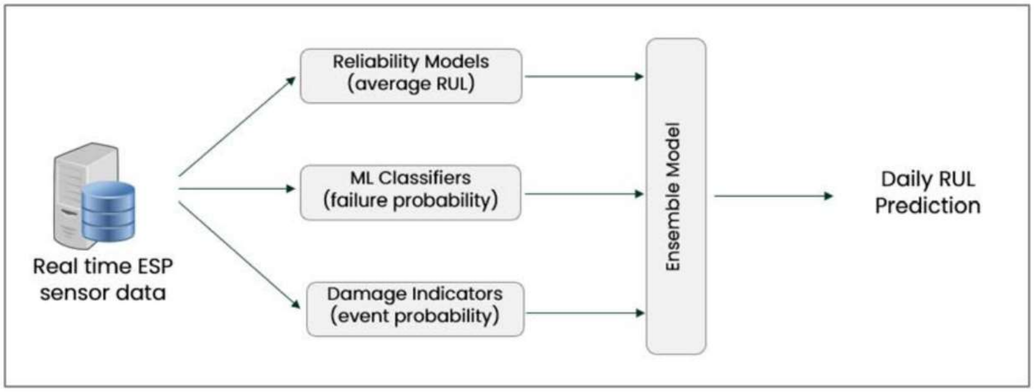


Figure 3—High level diagram of PFA algorithm

Source: Silvia, Shejuti *et al.* (2022). "Case Study: Predicting Electrical Submersible Pump Failures Using Artificial Intelligence." Paper presented at the Offshore Technology Conference, Houston, Texas, USA, May 2022. doi: https://doi.org/10.4043/31852-MS

11

Failure Prediction for ESP using ML – Case Study 2

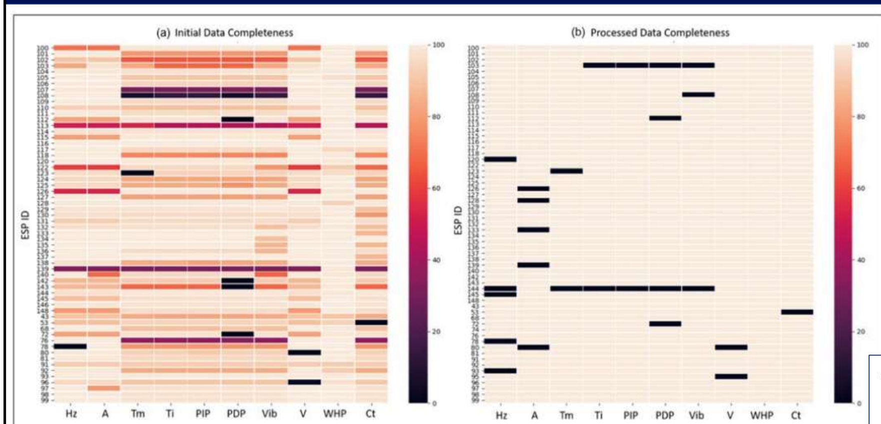


Figure 4—Data completeness comparison. (a) Data completeness prior to data interpolation. (b) Data completeness after data interpolation. Y-axis represents different ESPs in our dataset. Each row stands for a different ESP. X-axis has ten columns one for each sensor being evaluated. Data completeness is represented using a color scheme, the darker the cell is the more data is missing. More specifically, when all the data is available for a specific ESP and tag, the corresponding cell will be colored in light orange (has a value of 100% completeness) whereas when all the data is missing for a specific ESP and tag the corresponding cell will be colored in black (has a value of 0% completeness)

Table 2—Confusion matrix for PFA performance

		Actual	
		Pump Running	Pump Failed
Prediction	Pump Running	28 (TN)	9 (FN)
	Pump Failed	8 (FP)	9 (TP)

Source: Silvia, Shejuti *et al.* (2022). "Case Study: Predicting Electrical Submersible Pump Failures Using Artificial Intelligence." Paper presented at the Offshore Technology Conference, Houston, Texas, USA, May 2022. doi: https://doi.org/10.4043/31852-MS

12

Failure Prediction for ESP using ML – Case Study 2

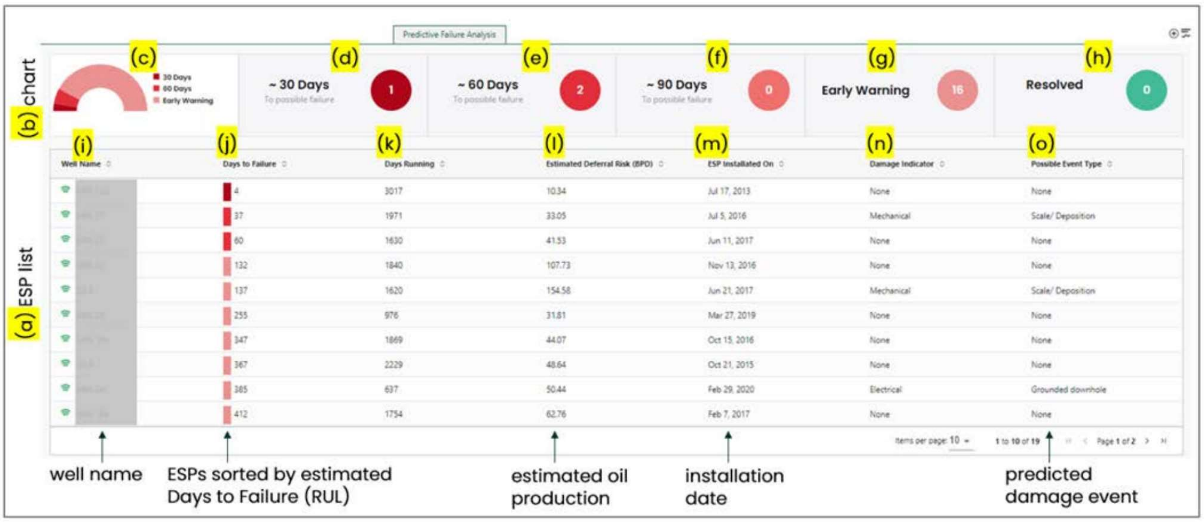


Figure 5—PFA Field View webpage for 36 running ESPs in the Americas fields

Source: Silvia, Shejuti *et al.* (2022). "Case Study: Predicting Electrical Submersible Pump Failures Using Artificial Intelligence." Paper presented at the Offshore Technology Conference, Houston, Texas, USA, May 2022. doi: https://doi.org/10.4043/31852-MS

13

Conclusions

- ML Solution paths are varied for ESP failure prediction.
- It requires more than real-time data sets.
- Not a single solution has been universally applicable.
  – You have to develop your own for your fields with our data.

14

Pan American ENERGY

8.0. Downhole Gauge Data – Reservoir Analysis

Based on the SPE Distinguished Lecture by Prof Roland Horne

Dr. Rajan Chokshi

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

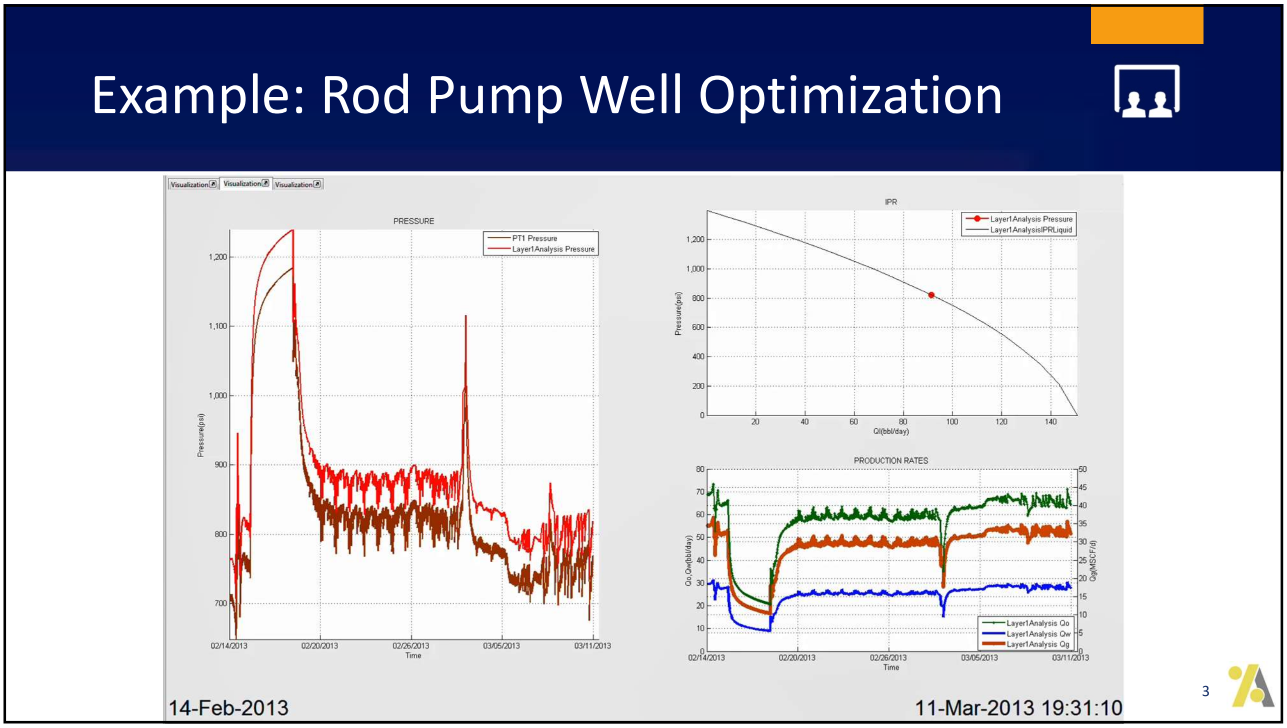
1

Permanent Downhole Electronic Gauges

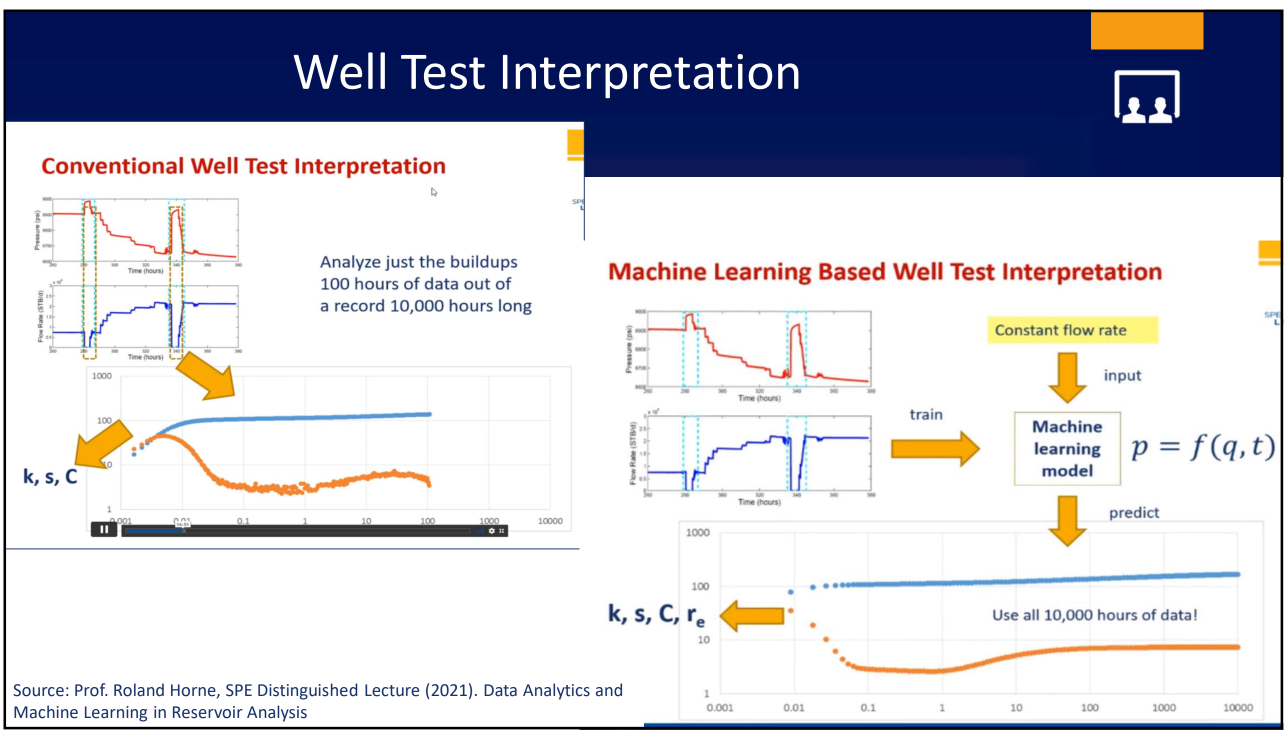
- Routine applications
  - Artificial lift monitoring, control & protection, producer-injector surveillance, zonal production management
  - P, T, Vibration measurements
    - Piezo-resistive strain up to 302°F and 10,000 psi
    - Quartz up to 392°F and 25,000 psi
    - Up to 16 gauges on a single conductor cable
- Data rarely applied for reservoir analysis

Source: SPE 176233, Gonzalez, Chokshi, & Lane, Real-Time Surface and Downhole Measurements and Analysis for Optimizing Production, doi:10.2118/176233-MS

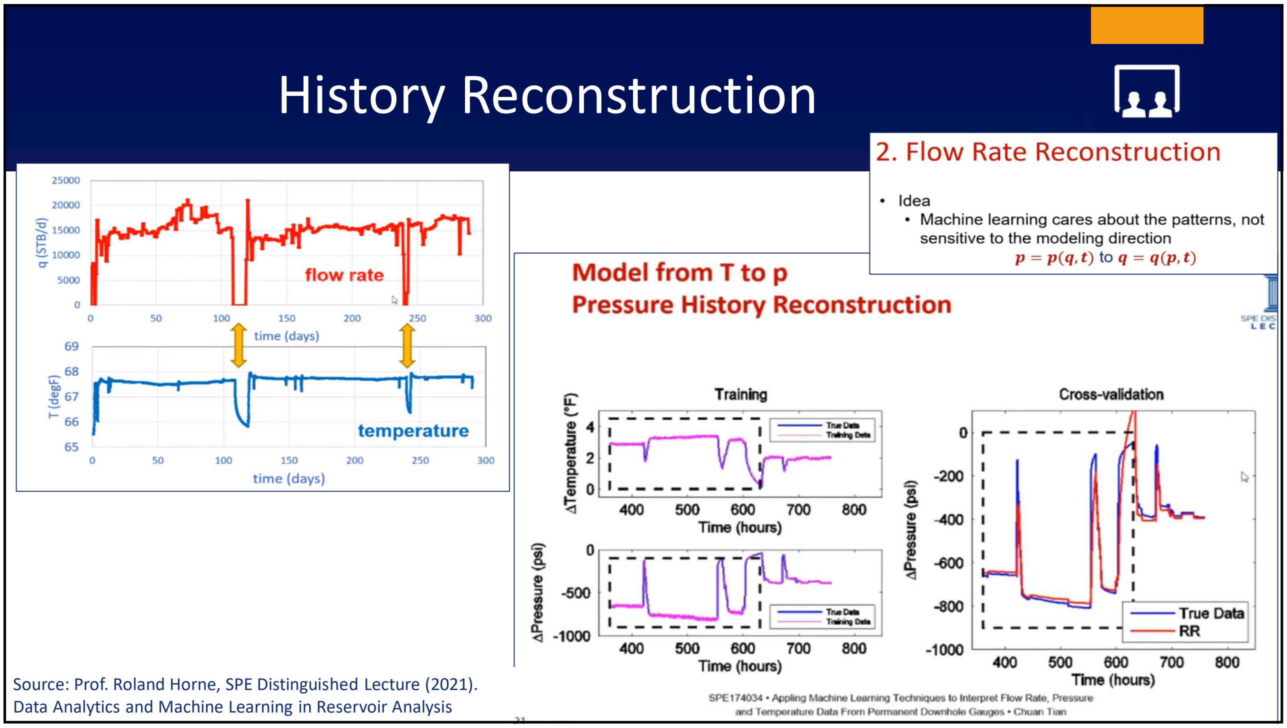
2


Example: Rod Pump Well Optimization
PRESSURE
PT1 Pressure
Layer1Analysis Pressure
IPR
Layer1Analysis Pressure
Layer1AnalysisIPRLiquid
PRODUCTION RATES
Layer1Analysis Qo
Layer1Analysis Qw
Layer1Analysis Qg
14-Feb-2013
11-Mar-2013 19:31:10
3

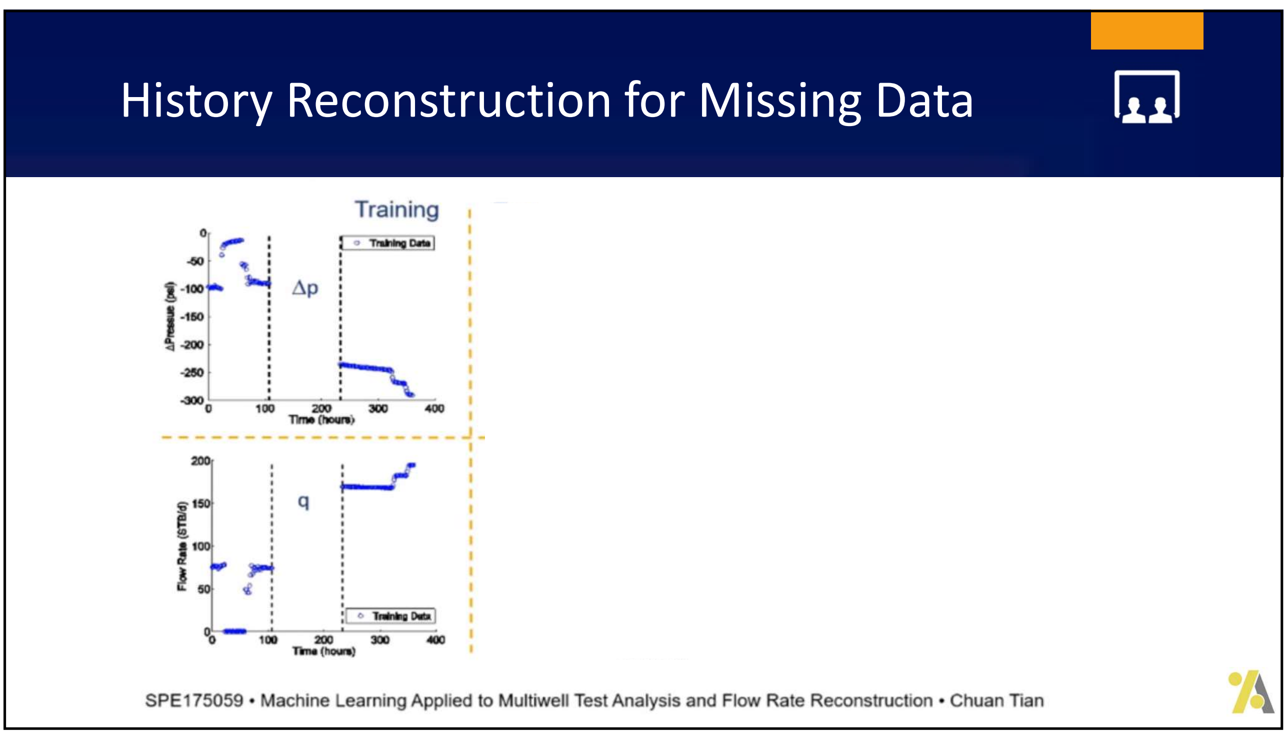
3


Well Test Interpretation
Conventional Well Test Interpretation
Analyze just the buildups
100 hours of data out of
a record 10,000 hours long
k, s, C
Machine Learning Based Well Test Interpretation
Constant flow rate
input
train
Machine
learning
model
$p = f(q, t)$
predict
k, s, C, r_e
Use all 10,000 hours of data!
Source: Prof. Roland Horne, SPE Distinguished Lecture (2021). Data Analytics and
Machine Learning in Reservoir Analysis

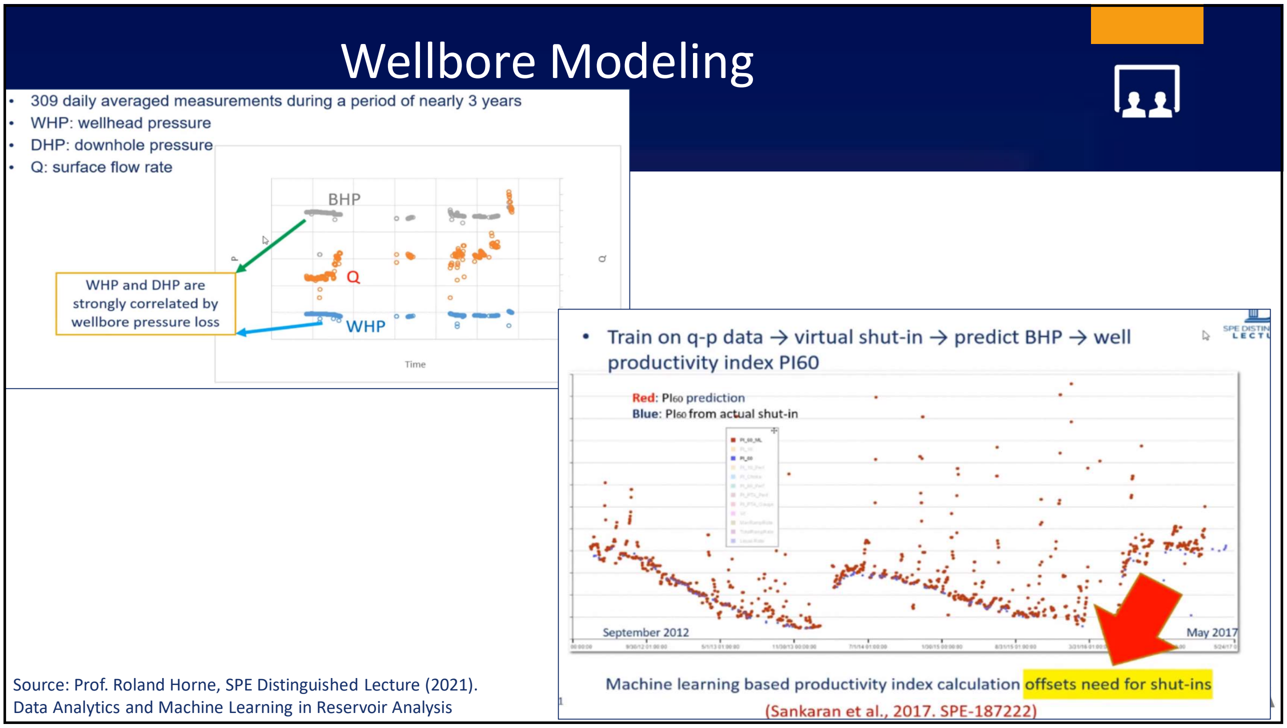
4


History Reconstruction
2. Flow Rate Reconstruction
• Idea
• Machine learning cares about the patterns, not sensitive to the modeling direction
$p = p(q, t)$ to $q = q(p, t)$
flow rate
temperature
q (STB/d)
T (degF)
time (days)
Model from T to p
Pressure History Reconstruction
Training
Cross-validation
ΔTemperature (°F)
ΔPressure (psi)
Time (hours)
True Data
Training Data
RR
Source: Prof. Roland Horne, SPE Distinguished Lecture (2021).
Data Analytics and Machine Learning in Reservoir Analysis
SPE174034 • Appling Machine Learning Techniques to Interpret Flow Rate, Pressure and Temperature Data From Permanent Downhole Gauges • Chuan Tian

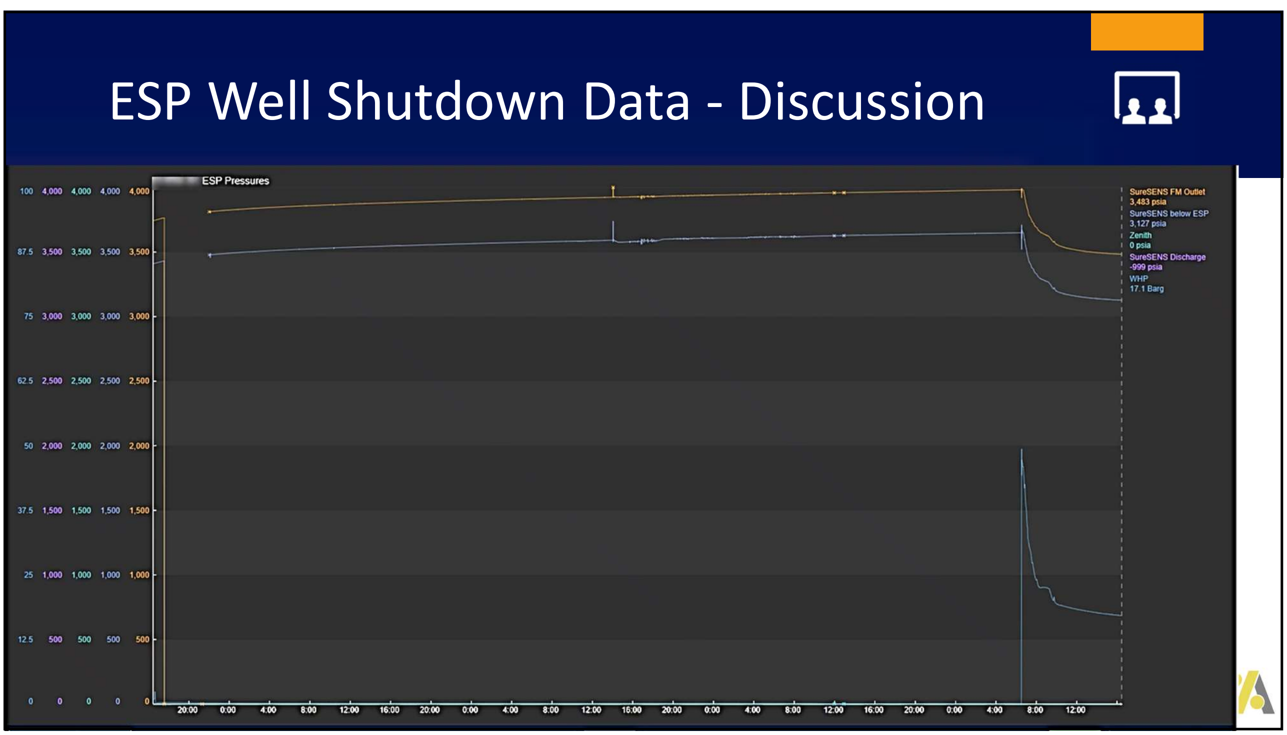
5


History Reconstruction for Missing Data
Training
Δp
q
ΔPressure (psi)
Flow Rate (STB/d)
Time (hours)
Training Data
SPE175059 • Machine Learning Applied to Multiwell Test Analysis and Flow Rate Reconstruction • Chuan Tian

6

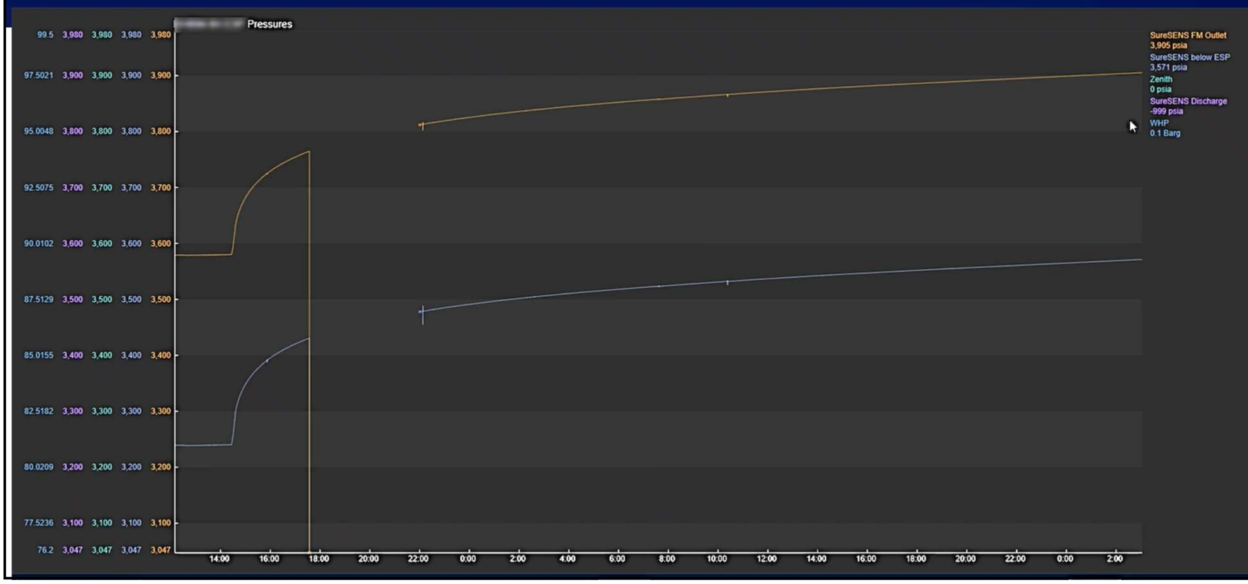

Wellbore Modeling
• 309 daily averaged measurements during a period of nearly 3 years
• WHP: wellhead pressure
• DHP: downhole pressure
• Q: surface flow rate
BHP
Q
WHP
WHP and DHP are strongly correlated by wellbore pressure loss
Time
• Train on q-p data → virtual shut-in → predict BHP → well productivity index PI60
Red: PI60 prediction
Blue: PI60 from actual shut-in
September 2012
May 2017
Machine learning based productivity index calculation offsets need for shut-ins
(Sankaran et al., 2017. SPE-187222)
Source: Prof. Roland Horne, SPE Distinguished Lecture (2021).
Data Analytics and Machine Learning in Reservoir Analysis

7


ESP Well Shutdown Data - Discussion
ESP Pressures
SureSENS FM Outlet
3,483 psia
SureSENS below ESP
3,127 psia
Zenith
0 psia
SureSENS Discharge
-999 psia
WHP
17.1 Barg

8

ESP Well Shutdown Data - Discussion



9

Overall

1. Permanent downhole gauges (PDG) **[q → p]**
   - Reveal the underlying reservoir model, *robust against noise, outliers and aberrations*!
   - Can use temperature data to add information and/or to substitute flow rate data.
2. Flow monitoring **[p → q]**
   - Can fill the unrecorded gaps in flow rate records.
   - Can allocate flow well-by-well based on manifold record.
   - Testing without shut-in!
3. Wellbore modeling **[WHP → DHP]**
   - Discovery of relationship between downhole and surface conditions can provide missing data or alert out of pattern events.

July 15, 2021 23

10

9.0. Optimal Gas Injection in a Single Point Gas lift for Gas Well Deliquification

Dr. Vinicius Kramer Scariot, Dr. Rajan Chokshi
Data & Script Contribution: Blazej Ksiazek and Antonio Reinoso
Advisor: Dr. Eduardo Pereyra, TUHWALP

Data Analytics for Artificial Lift & Production Optimization

1

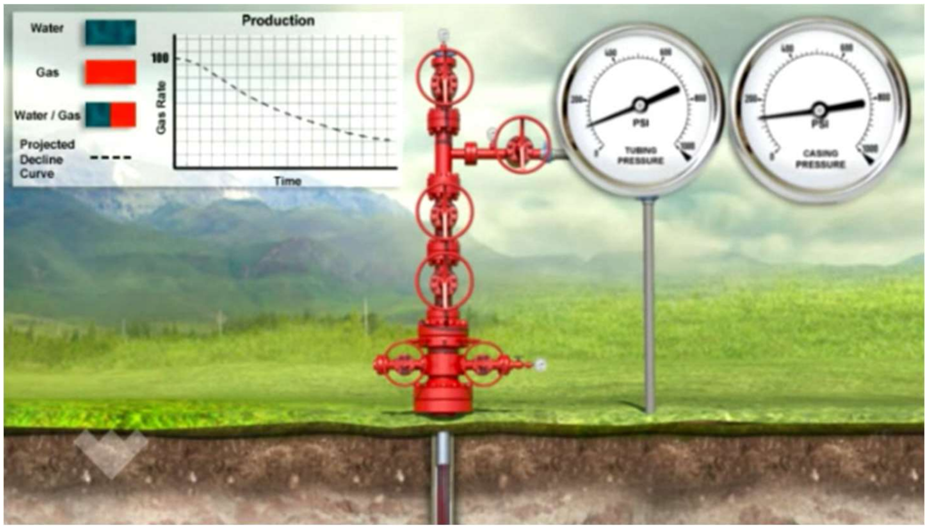
Outline

- Introduction
- Objective
- Data gathering
- Neural Network Design
- Proper Implementation
- Summary
- Takeaways

2

2

Liquid Loading Animation

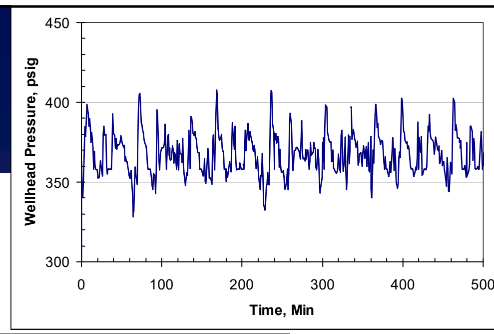


Courtesy: Weatherford

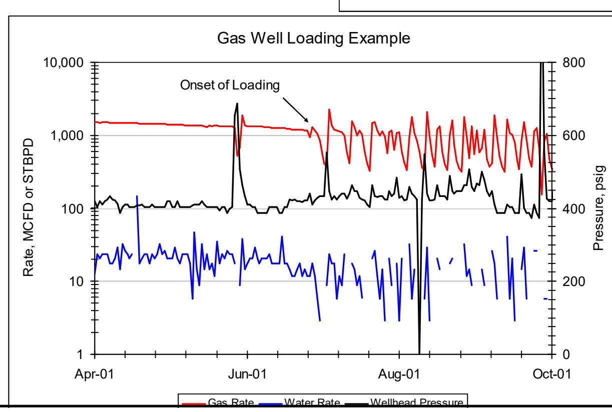
3

3

Common Signs of Liquid Loading

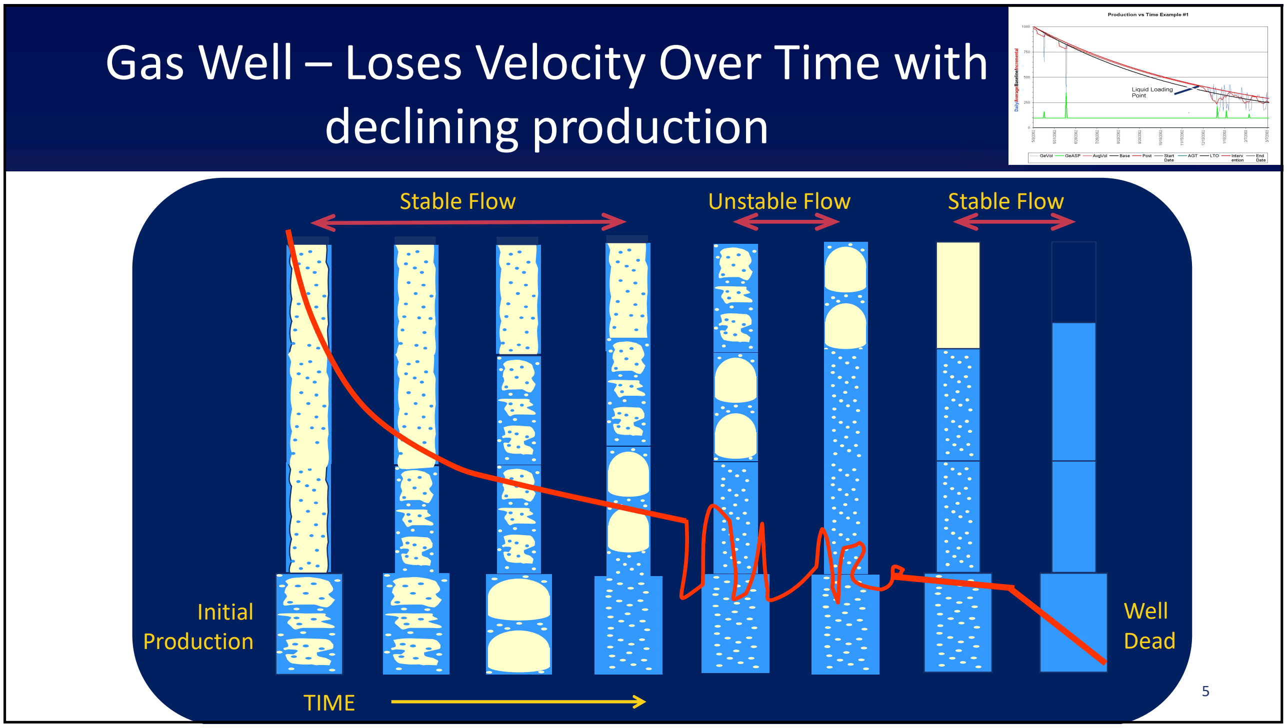


- Tubing and casing pressure differential (packerless completion)
- Pressure spikes
- Liquid slugging
- Fluctuating gas production
- Liquid production stops all together

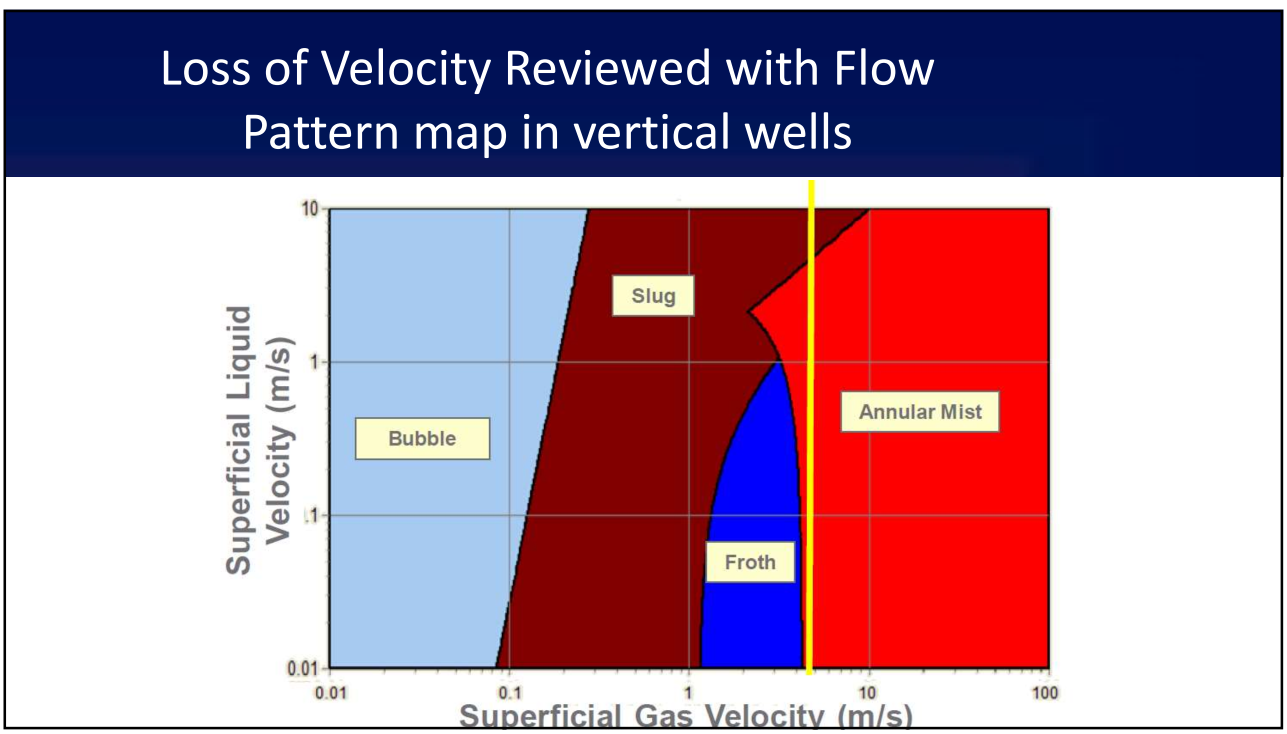


4

4


Gas Well – Loses Velocity Over Time with declining production
Stable Flow
Unstable Flow
Stable Flow
Initial Production
Well Dead
TIME
5

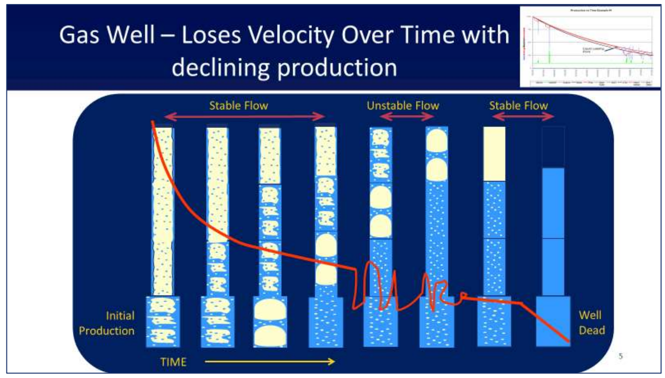
5


Loss of Velocity Reviewed with Flow Pattern map in vertical wells
Superficial Liquid Velocity (m/s)
Superficial Gas Velocity (m/s)
Bubble
Slug
Froth
Annular Mist
10
1
.1
0.01
0.01
0.1
1
10
100

6

Liquid Loading Cause & Remedy

- Liquid Loading:
  - "Inability of the produced gas to remove the produced liquids from the wellbore"
  - Decrease in gas velocity.
  - Accumulation of liquids at the bottomhole.
- Artificial-lift method required to continue production.
  - In this application we will consider gas-lift to remove liquid from the wellbore.



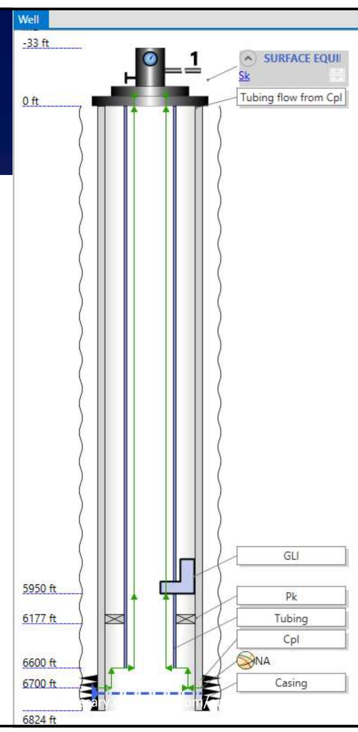
Riza, M. F., Hasan, A. R., and C. S. Kabir. "A Pragmatic Approach To Understanding Liquid Loading in Gas Wells." *SPE Prod & Oper* 31 (2016): 185–196. doi: **https://doi.org/10.2118/170583-PA**

7

7

Gas Lift

- What is gas lift?
  - Reduce bottomhole pressure
  - Valve position
  - Operating pressure
  - Gas injection rate

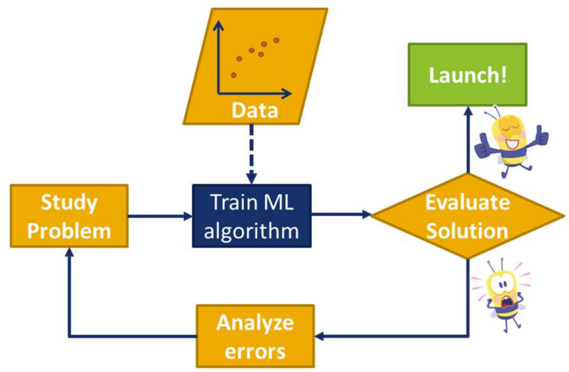


8

8

Objective of this study

- Using a machine learning approach, find the optimum amount of gas to be injected in a single point gas lift in order to maximize production at a specified reservoir pressure.
- Must have available data for the well.
- In this study, we will use
  - A simulation data set, prepared using PipeSim software for a gas-lift well, will be used to train and test a Neural Network Model.
- Would you call this a **hybrid** model?!?

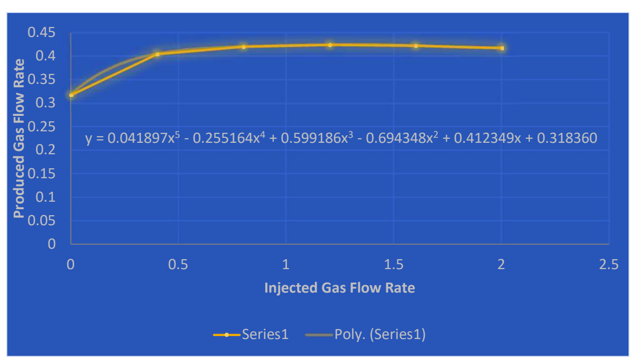


9

9

Data Generation

- For each reservoir pressure:
  - Several production flow rates can be predicted for different injection rates
  - Based on the calculated production rate, the optimal injection rate can be predicted.
  - For the next reservoir pressure, the backpressure model can be written as: $C_{Future} = C_{Present} \times \dfrac{\bar{p}_{res(future)}}{\bar{p}_{res(present)}}$



10

10

Data Generation Simulations

- Well specifications
- Optimized Variable:
  - Gas Produced
- Independent Variables:
  - Gas injected
  - Reservoir Pressure

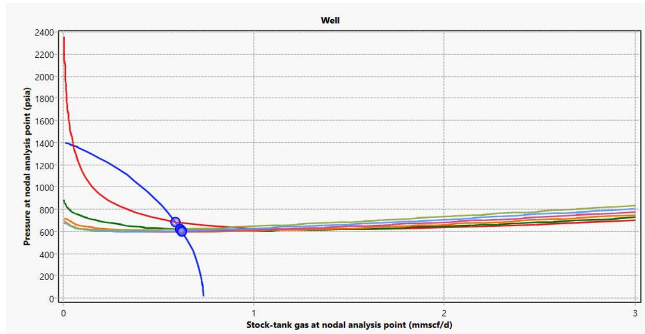
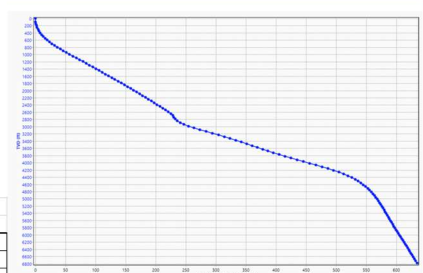


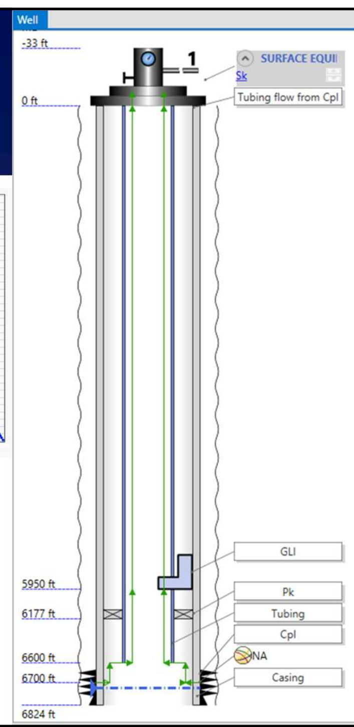
Figure 1. IPR & OPR curve at 1400 psi

11

11

Well Specifications





Well Type	Production			
Check Valve	Block reverse			
Tubulars	**Bottom MD**	**ID**	**Wall thickness**	**Roughness**
Casing	6824	4.778	0.361	0.001
Tubing	6600	1.61	0.145	0.001
Downhole Equipment	**MD**		**Viscosity**	
Packer	6500		Under sat	Vasquez & Beggs
Artificial Lift	MD		Live oil	Chew & Connally
GLI	6450		Dead Oil	Beggs & Robinson
Gas sg	0.6		T1	200
Completions			V1	0.902
Single point	6700		T2	60
Fluid			V2	11.91
WGR	12	stb/mmscf	**Water Cut Inversion**	60%
OGR	0	stb/mmscf		
Gas sg	0.6636			
Water sg	1.1			
API	45			

	Specific heat capacity		Thermal conductivity	
Gas	0.5500017	Btu/(lbm.degF)	0.02	Btu/(h.degF.ft)
Oil	0.4500014	Btu/(lbm.degF)	0.08	Btu/(h.degF.ft)
Water	1.000003	Btu/(lbm.degF)	0.35	Btu/(h.degF.ft)

Enthalpy calculation method: 1983 2009

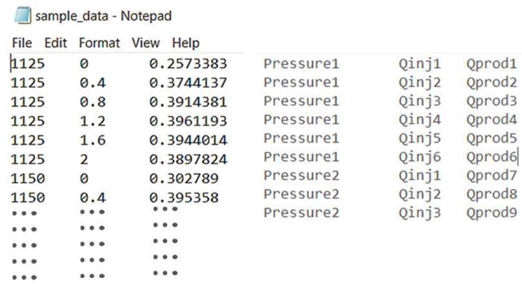
Specific latent heat of vaporization: 139.9996 Btu/lbm

12

Data Organization

- Inputs : Sample_data.txt

- Code divides the input data into training and testing subsets.


sample_data - Notepad

File Edit Format View Help

1125	0	0.2573383	Pressure1	Qinj1	Qprod1
1125	0.4	0.3744137	Pressure1	Qinj2	Qprod2
1125	0.8	0.3914381	Pressure1	Qinj3	Qprod3
1125	1.2	0.3961193	Pressure1	Qinj4	Qprod4
1125	1.6	0.3944014	Pressure1	Qinj5	Qprod5
1125	2	0.3897824	Pressure1	Qinj6	Qprod6
1150	0	0.302789	Pressure2	Qinj1	Qprod7
1150	0.4	0.395358	Pressure2	Qinj2	Qprod8
...	...	...	Pressure2	Qinj3	Qprod9
...	...	...			
...	...	...			
...	...	...			
...	...	...			

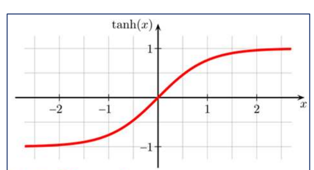
13

13

Neural Network Development in Python

- Libraries used:
  - Numpy
  - Pandas
  - Matplotlib
  - Sklearn
  - Tensorflow

- Code Specifications:
  - Two Inputs: Reservoir Pressure, Injection Rate
  - One Output: Production Rates
  - Two Hidden Layers with four neurons each
  - Tangent Hyperbolic **Activation Function: tanh**

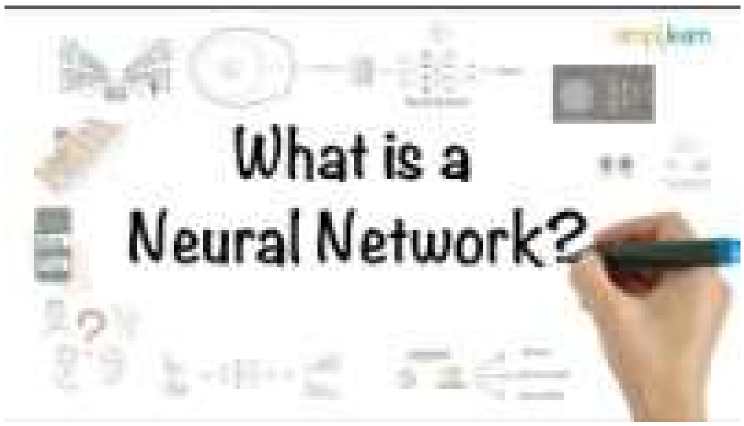


An activation function outputs a small value for small inputs, and a larger value if its inputs exceed a threshold. If the inputs are large enough, the activation function "fires", otherwise it does nothing. In other words, an activation function is like a gate that checks that an incoming value is greater than a critical number.

  - 200 **Epochs**
    - an epoch refers to one cycle through the full training dataset. Usually, training a neural network takes more than a few epochs.

14

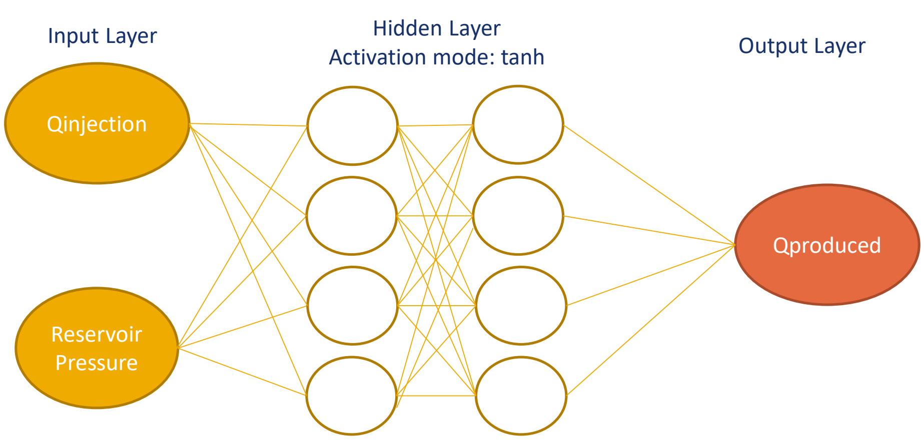
Neural Network – Explanation



Source: Neural Network In 5 Minutes | What Is A Neural Network? | How Neural Networks Work | Simplilearn, https://www.youtube.com/watch?v=bfmFfD2RIcg

15

Neural Network Model



16

16

Summary

- Possible applications:
  - Analyze multiple wells quickly.
  - Combine with simulation software.
  - Track reservoir pressure in real time.

Production decreases due to liquid loading

Record reservoir pressure

Run program to find optimum injection rate

Implement solution

17

17

Takeaways

- This project explored and proved the feasibility of machine learning as a tool to solve industry production problems.
- The main benefit of using machine learning is that it is capable of solving problems involving large data sets in short amount of time and accurately.

18

18

Pan American ENERGY

10.0. Slug Catcher Design: Flow Patterns Application

Dr. Vinicius Kramer Scariot

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

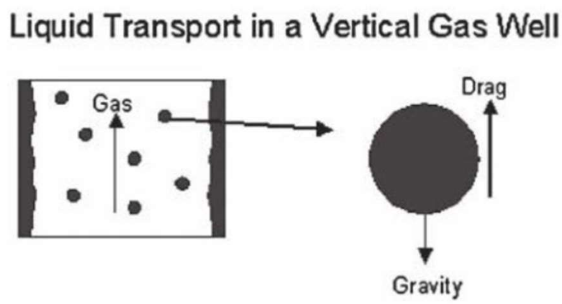
Outline

- **Applications**
- Gas-Liquid flow patterns
- Data analytics
  - Database
  - Neural networks

2

Applications of Flow Pattern Prediction

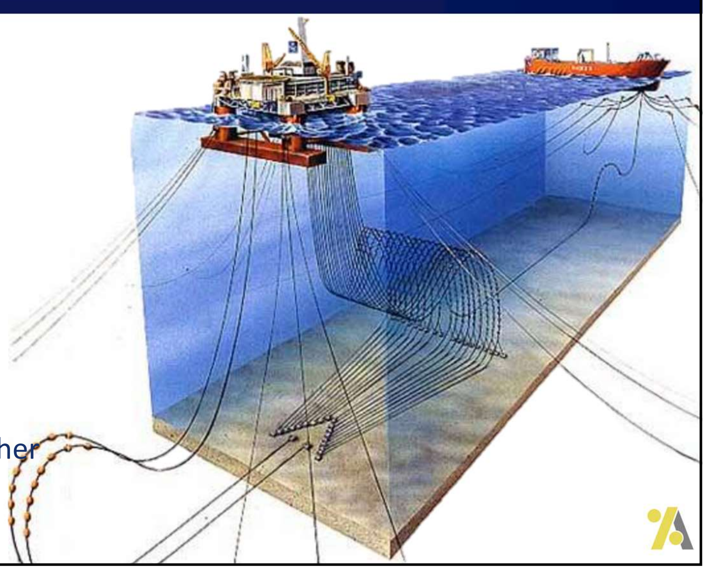
- Liquid loading
  - At late production, gas velocity is not sufficient to lift liquids
    - Formation water
    - Condensed hydrocarbons
  - Liquid accumulation
    - Reduced flow rates
    - Killing of well
  - Deviation from annular flow



3

Importance of Flow Pattern Prediction

- Severe slugging
  - Liquid overflow
  - Separator shutdown
  - Problems in flaring
  - Erosion in riser
- Detection
  - Stratified flow in pipeline
    - Bøe Criterion to evaluate further

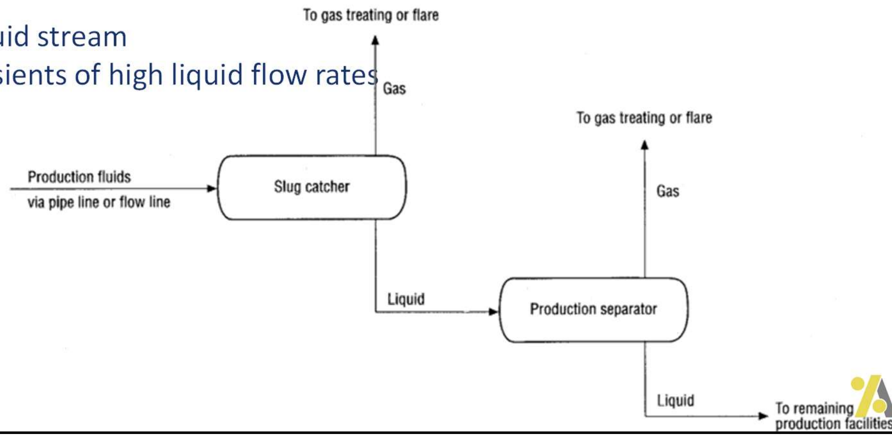


4

Importance of Flow Pattern Prediction

Slug catchers are Process equipment that averages the flow transient behavior

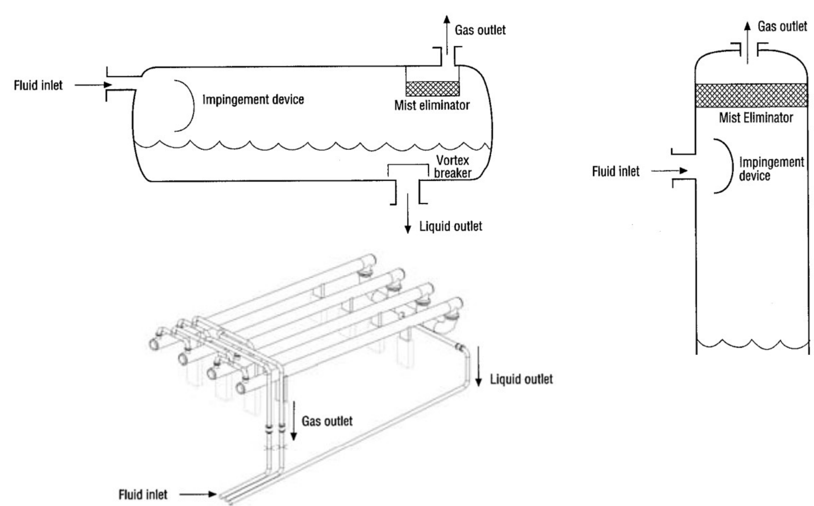
- Enough capacity
- Steady outlet liquid stream
- Attune inlet transients of high liquid flow rates



5

Types of Slug Catchers

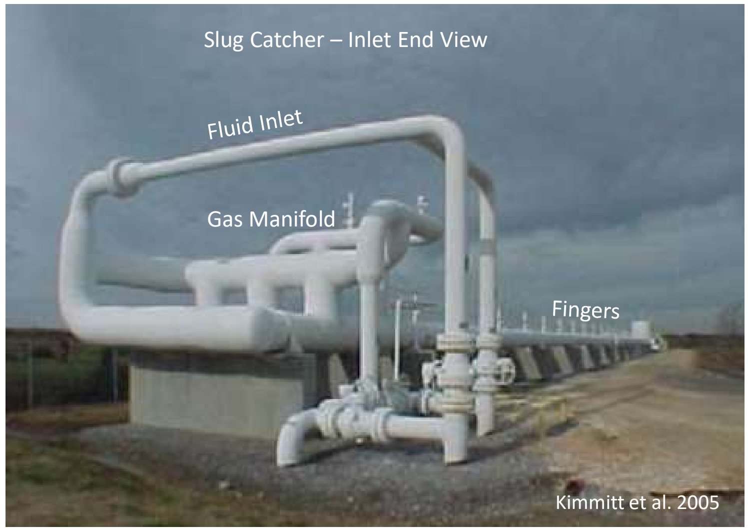
- Horizontal
- Vertical
- Finger type



6

Finger Type Operation

- Advantages
  - Piping components
- Finger operation
  - Stratified flow
- Finger design
  - Capacity
  - Flow pattern



7

Outline

- Applications
- **Gas-Liquid flow patterns**
- Data analytics
  - Database
  - Neural networks

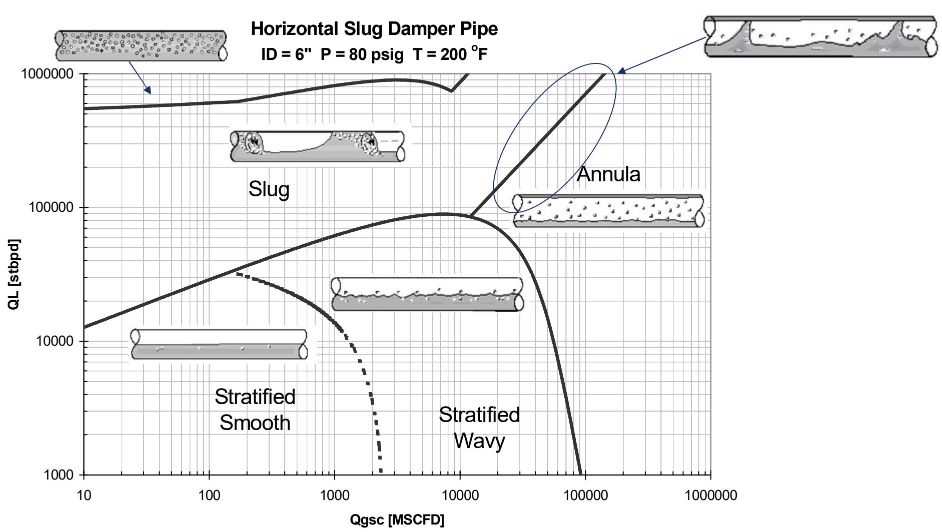
8

Flow Patterns Two Phase Flow

Each phase in a gas-liquid flow has a flowing resistance which depends on thermo-physical properties (μ_i, ρ_i, σ) and flow conditions ($v_{s,i}, \theta, D$, etc.). The resulting phase distribution for given conditions is known as a flow pattern.

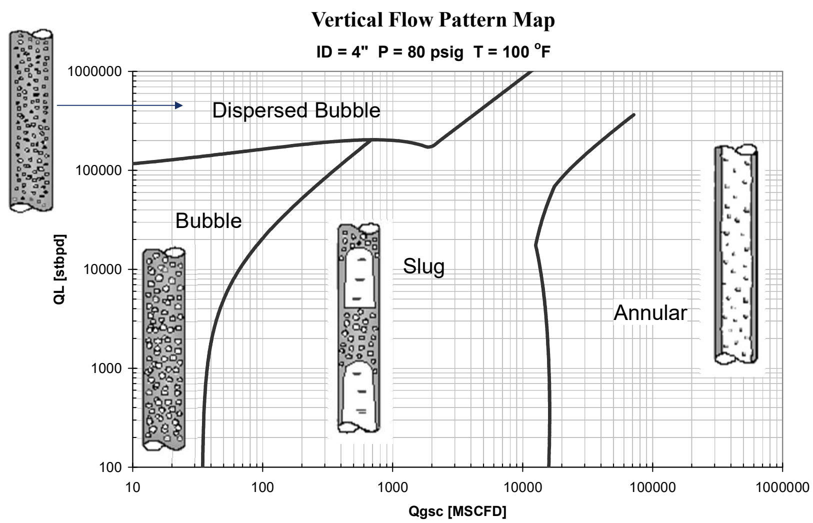
9

Horizontal Flow Patterns



10

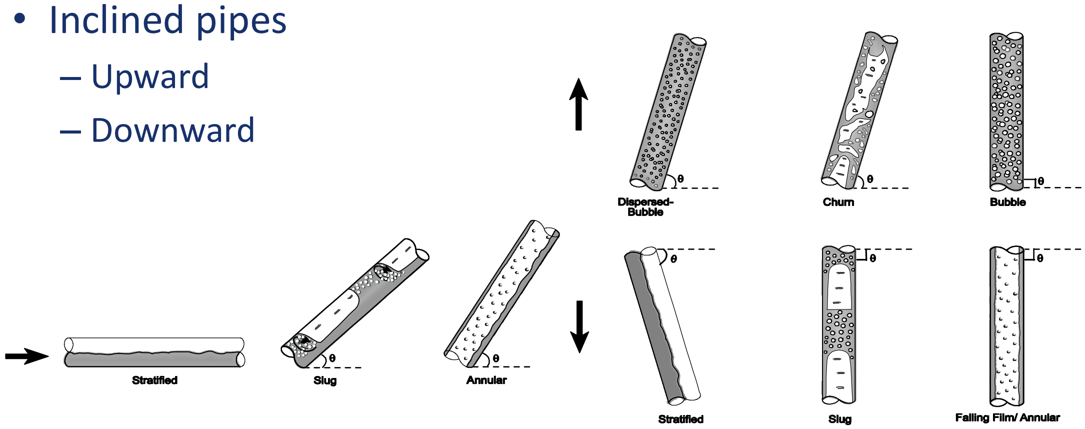
Vertical Flow Patterns



11

Inclined Flow Patterns

- Inclined pipes
  - Upward
  - Downward



12

Flow Pattern Models

Physics model (e.g. Barnea)

- Simplified Physical Mechanisms
  - Buoyancy vs. Turbulence
  - Instability growth
  - Bubble characterization (rise velocity, size)
- Use of data-driven correlations for closure relationships
- Very well explored
  - Limitations are known
  - Vast literature

Data driven model (e.g. Neural Network)

- Reliance on the training data
  - Validity
  - Proximity
- Hard to interpret results
  - Verify
  - Lack of insights
- Just started being explored
  - Lack of confidence
  - Use with caution

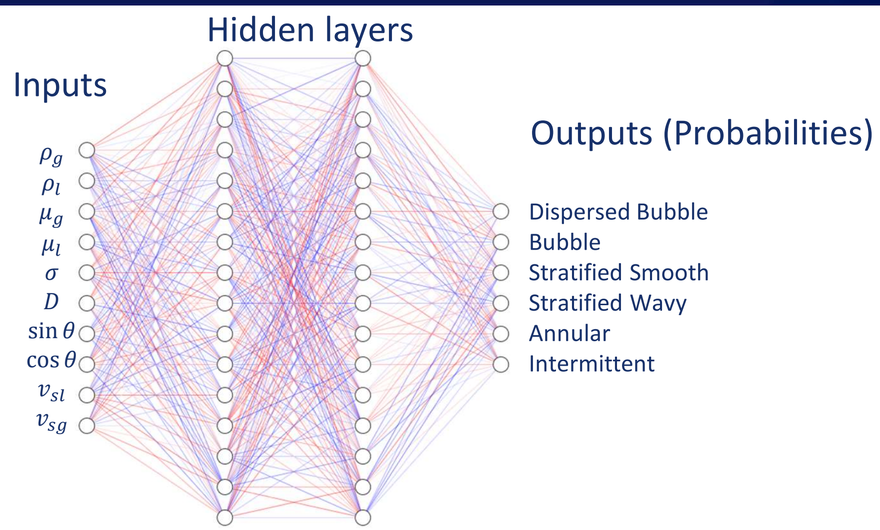
13

Outline

- Applications
- Gas-Liquid flow patterns
- **Data analytics**
  - Database
  - Neural networks

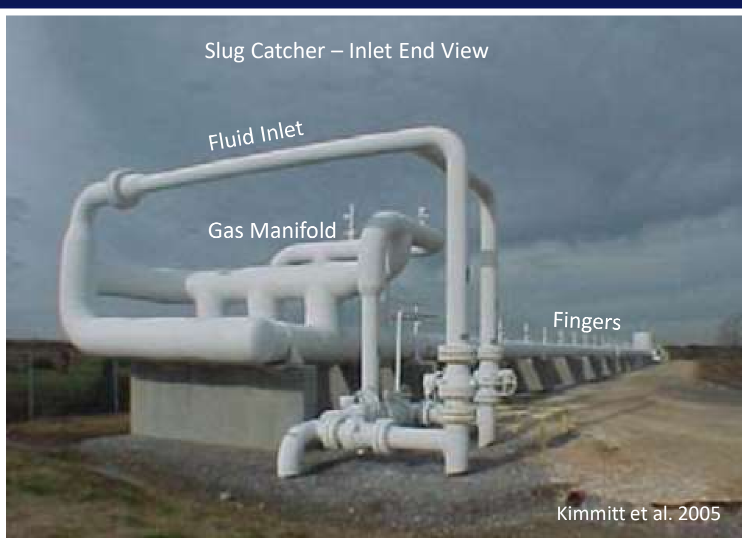
14

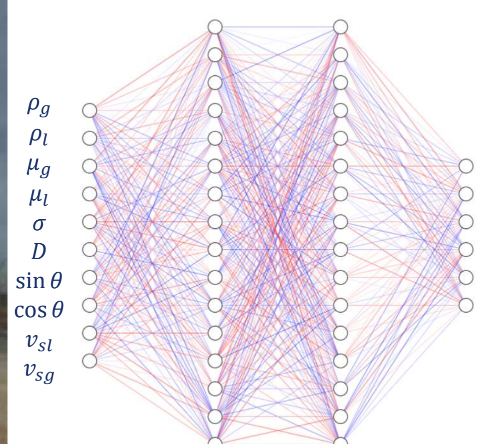
Neural Networks



15

Application – Finger Slug Catcher Design





16

Pan American ENERGY

11.0 Network Optimization Using Neural Networks and Genetic Algorithms – SPE 201760

Sandip Melkaveri Dr. Rajan Chokshi

Presentation at ALRDC Gas Lift Workshop Training Course Jun 2021

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

Introduction

Entire presentation is based on a recent publication.

SPE-201760-MS: Network Optimization Models at Greater Kuparuk Area Using Neural Networks and Genetic Algorithms.

Authors: Murray, Rodney L., Hopkins, Reese S., and Douglas K. Valentine.

Paper presented at the SPE Annual Technical Conference and Exhibition, Virtual, October 2020. https://doi.org/10.2118/201760-MS

This reference is selected because of comprehensive treatment of network model applied on a sizeable asset, and recent publication-timeline. All figures, tables in this presentation are from the above reference unless noted otherwise. No rights are claimed.

2

Area of study – Greater Kuparuk Area (GKA), North Slope, Alaska

1981 | Year of first production

2.5 B | Total bbls produced till date

> 1200 | Total wells (producers + injectors)

90 | % of wells on gas lift in GKA

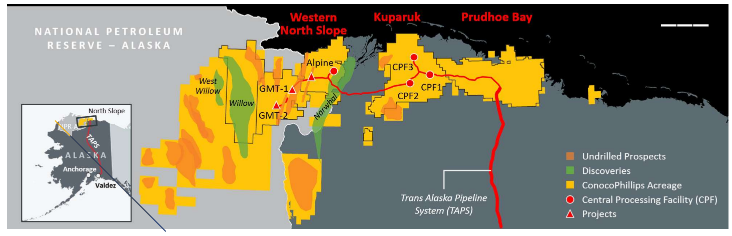


Image Courtesy: Google

3

Objective

- Develop a fast, flexible optimization model that recommends well status, lift gas rates, and water injection rates [*sic*]
- Data used in the building of model includes
  - Field data
  - Data generated by previous surface models in the development of hydraulic models
  - Current facility conditions
  - Constraints

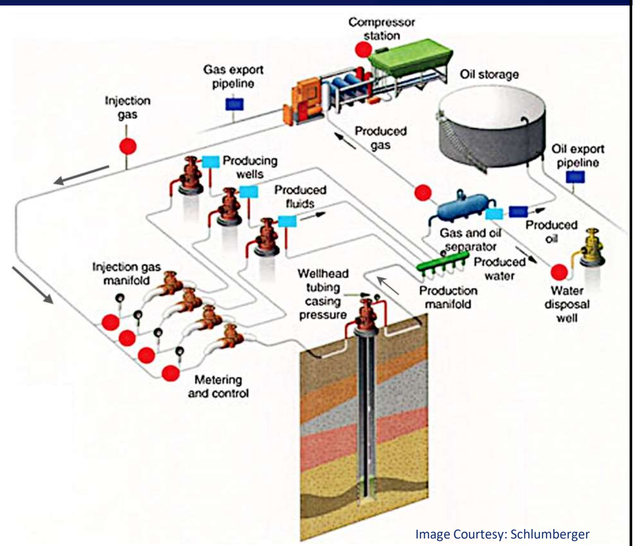


Image Courtesy: Schlumberger

4

Need for optimization

- The facilities are capacity constrained
  - the injection gas lift compressors, and
  - the injection pump for water disposal
- A previous attempt to more rigorously optimize the production system using commercial software resulted in better lift gas allocation, but computation time led to the cessation of its use for daily optimization
- An optimization program using the equal slope concept is currently in use for lift gas allocation (drawback - does not consider back-pressure on the entire system)
- GKA Network Optimization Model (GNOME) - solution for production and injection network optimization problems

5

Previous efforts

- Commercial software
  - Physical equations and correlations to solve back-pressure
  - Time consuming for model to converge
- 6-approaches method by Rashid et al
  - Rashid, K., Bailey, W., & Couët, B. (2012). "A survey of methods for gas-lift optimization. *Modelling and Simulation in Engineering*, 2012".
- Use of synthetic data to train neural networks
  - Shokir, Hamed, Ibrahim, & Mahgoub, 2017. "Gas Lift Optimization Using Artificial Neural Network and Integrated Production Modeling. *Energy and Fuels*, **31**(9), 9302–9307".

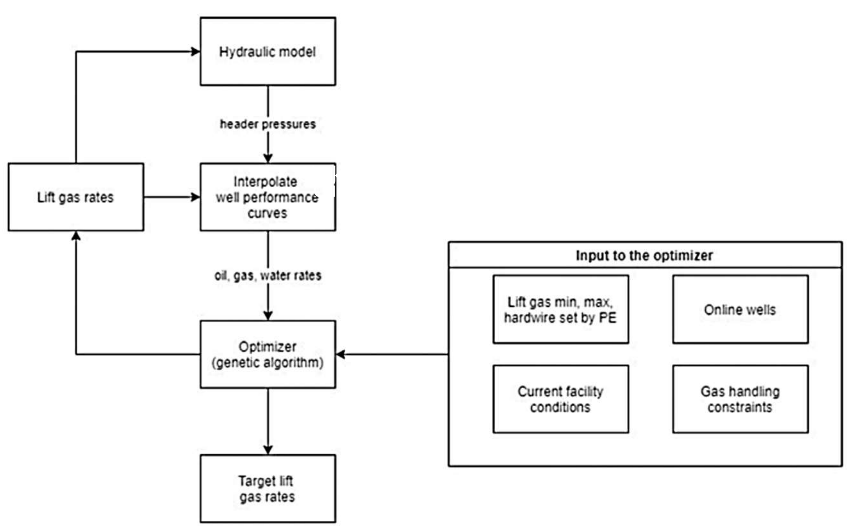
6

Model framework

- Component 1: A function that estimates producer and injector performance
- Component 2: A function that gathers and interpolates well performance models with physics-based models
- Component 3: Estimating drillsite header pressures using a neural network
- Component 4: Genetic algorithm used for searching the optimal well status, lift gas rate, and water injection rate for each well

7

Process flow diagram – production network



8

Process outline

- Step 1: Identify facility constraints and evaluate well performance models
- Step 2: Hydraulic model to determine pressure at drillsite
- Step 3: Genetic algorithm to solve for lift gas recommendations

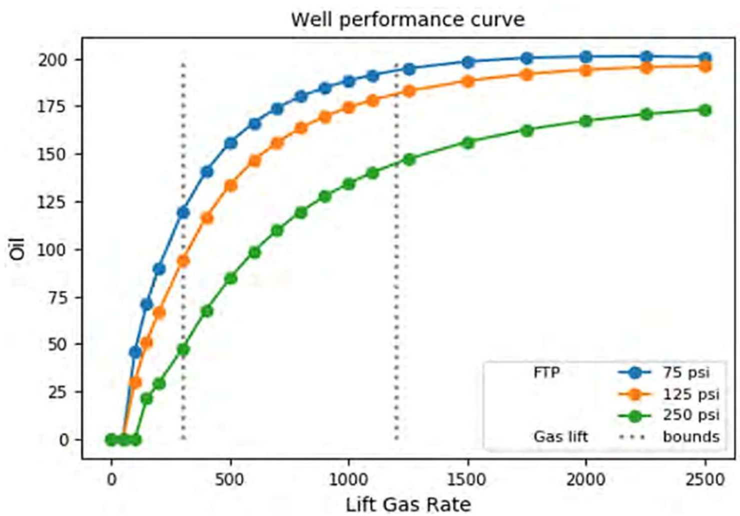
9

Step 1: Create scenario and run well performance evaluation

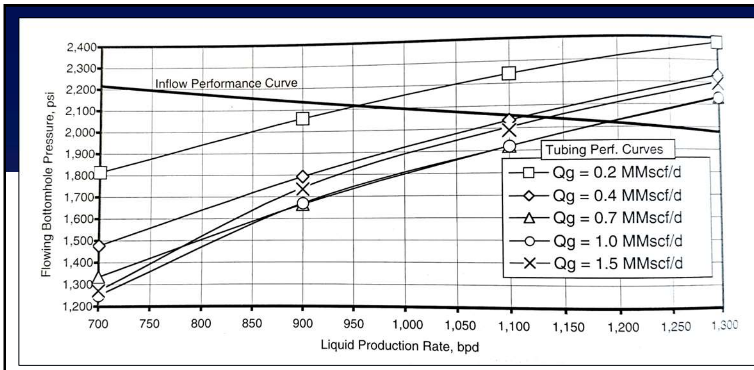
- Scenario to be loaded into the program
- Static data
  - Res Pressure
  - Completion schematic
- Dynamic data
  - WHP, liquid rates
  - Casing pressure

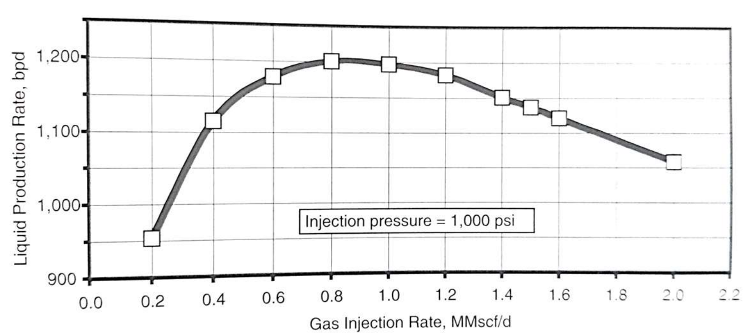
10

Well performance evaluation



11





Picture Courtesy:
Gas Lift Manual - Gabor Takacs

12

Step 2: Create a hydraulic model

- Hydraulic model estimates pressure drops in pipeline. It is trained in Neural Network using a two-step approach
  - Step 1: Generate random, uniformly distributed O/G/W rates
  - Step 2: Model was trained on 5 years of data including one-hour averages of rates

Input nodes	Output nodes
O/G/W rates, WCut, GLR, Sep inlet press.	Drillsite header pressure

End Result: GNOME performed on par with the other network simulation models

13

Hydraulic model results comparision

	CPF1	CPF2	CPF3
Network simulation	10.4 psi	6.4 psi	4.5 psi
Neural network	6.3 psi	6.5 psi	5.5 psi
Number of hydraulic checks	158	165	206
Number of drillsites	13	20	15

14

Water injection model

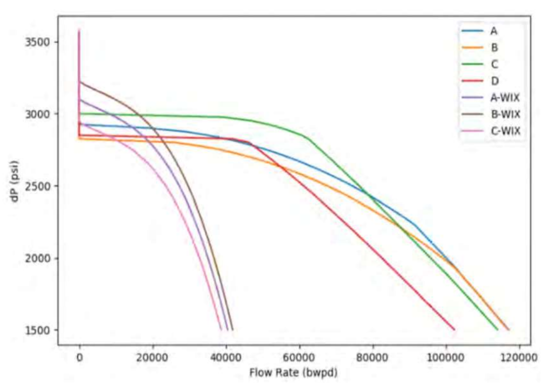
- The model is solved at well level in each facility
- Injection rates are calculated based on injection pressure and constraints
- Injectivity index (the equivalent of PI in a producing well) is the defining metric
- Min WHP, max WHP and max rate are defined

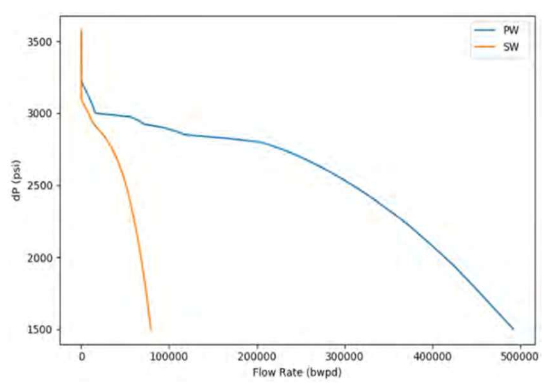
$$\mathrm{VRR} = \left(\frac{\mathrm{Tot.\,Vol\ Injected}}{\mathrm{Tot.\,Vol\ Produced}}\right)_{res\ cond}$$

15

Water Injector Pump Performance

- Two different types of water – produced water and sea water
- Parameters were introduced to allow for tuning of the pump curves to match observed performance
- Curve fit algorithms are used to automatically adjust these parameters to actual field data





16

Pipeline network modeling

- Linear regression model was used to model the pressure changes
- Inputs
  - Injection rates for each of the drillsites
- Outputs
  - Δp from the CPF to drillsites
- Trained on 4 months of historical data
- Mean absolute error of 5 PSID

17

Solving the water injection network

- First, assume a constant pressure at the facility, ignoring any constraints of the pumps.
- Second, assume a constant discharge pressure from the pumps using pump curves. This translates to constant inlet pressure and assumes any unlimited amount of water for injection
- Lastly, the network is solved assuming that all water produced must be injected. Shut down wells or open them to their maximum capacity based on water availability

18

Step 3: Genetic Algorithm to solve for optimized solution

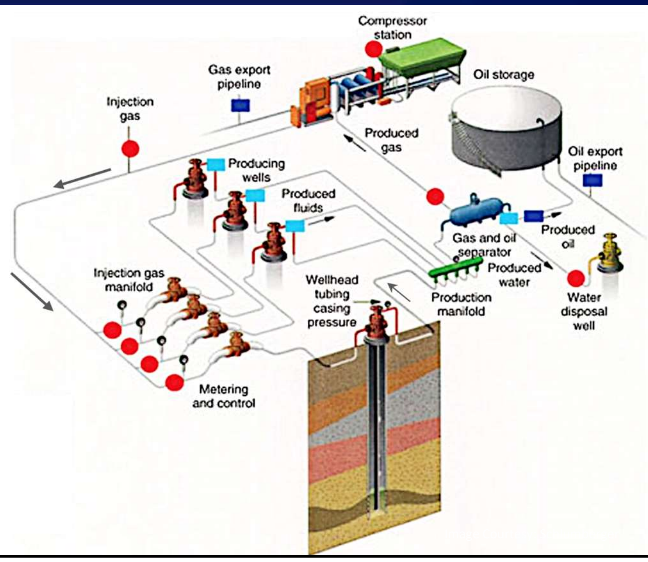
op·ti·mi·za·tion: [op-t*uh*-m*uh*-**zey**-sh*uh*n]

a mathematical technique for finding a maximum or minimum value of a function of several variables, subject to a set of constraints.

19

Gas lift optimization scenarios

- The simplest scenario…. is an individual stand-alone well.
- A complex scenario…. is a large system with numerous wells producing into a common gathering system.



20

Gas lift optimization – Controls and Constraints

- Objective
  - Maximize production/income
- Controlling parameters
  - Choke settings (down hole, wellheads and manifolds)
  - Flow control valve settings (equiv. area, max. rate through valve)
- Constraining parameters
  - compression capacity
  - max. GOR, water cut
  - max. liquid, max. Gas
  - (min) flow rate, available lift gas

21

GNOME

- Objectives – well status (S) & lift gas rate (L)
- Constraint – compression capacity (C)
- Genetic algorithm to solve for optimized L & S values
- Inspired by biology and evolution, GA works well with data that is discreet, discontinuous or noisy

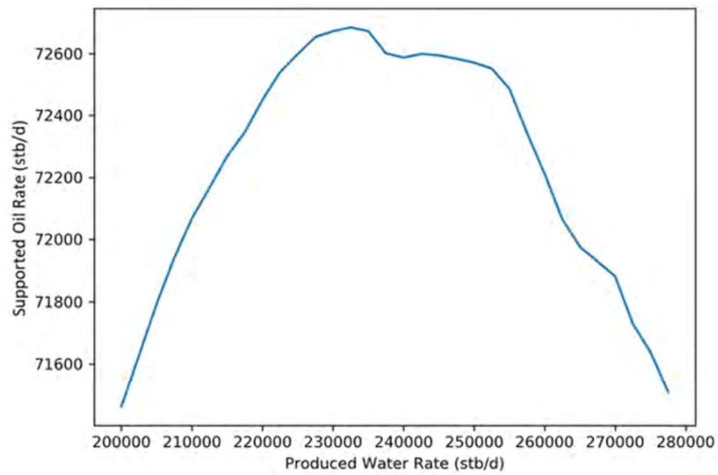
$$\sum_{i=1}^{N}(L_i + Q_{gi})S_i \leq C$$

Link to video on GA:
https://www.youtube.com/watch?v=3QJjfeVrut8

22

Combining Production and Water Injection Network Models

- Pre-calculate the relationship between total supported oil rate as a function of the total water production



23

Use Cases

- GNOME is run daily. It compares rates that were implemented yesterday to what GNOME recommended today.
- Evaluating the net benefit of a producing well that is online (run daily)
- Determining gas lift capacity sensitivity (run daily)
- Drillsite injection configuration changes

24

Summary

- GNOME combines physical equations and models with machine learning to deliver powerful optimization
- GNOME currently runs over 1000 optimizations daily with a different scenario
- Production network model based on well performance curves and 3-phase hydraulic model
- Injection network model is based on pump model and single-phase hydraulic model
- The recommendations from the optimization program are expected to increase oil rate 1.5% in the existing system

25

Pan American ENERGY

12.0. Multiphase Flow Meter Virtualization

Dr. Patrick Bangert
Algorithmica

Dr. Rajan Chokshi
Accutant Solutions LLC

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

Accutant Solutions
Accurate Accountable Acumen

1

Executive Summary

- This project is a proof of concept to determine if a **virtual multiphase flow meter** can be constructed using machine learning rather than physical modeling.
- For 23 wells, we had 10 tags each and **158 historical well tests** to supply training information.
- Models were constructed for **oil, water, and gas flows** and found to be accurate to within **15%, 8%, and 10%** respectively in all but exceptional well tests.
- These models are **easy to deploy and scale** with no additional effort. They can also be improved easily with fresh information in the future. Initial deployment would require some software development.

 2

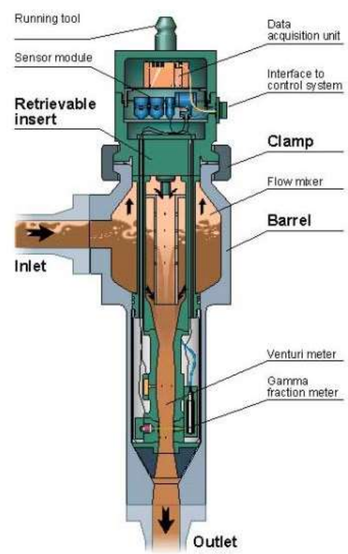
2

The Study

3

Multiphase Flow Metering

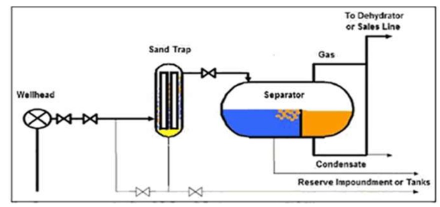
- Wells produce **multiphase fluid**: Water, oil, gas, and often particulate matter
- We can easily measure the flow rate of the fluid, pressures and temperatures around the well.
- We want to measure how much oil, gas, and water is in the fluid – Multiphase Flow Meters
- Meters are **expensive and fragile** – we want virtual meters that **compute instead of measure**



4

Well Testing

- Wells can be connected to a separator that physically separates the three phases
- Each phase can then be measured
- This is done at xxx and we have the data from these well tests
- The well test data forms the baseline for our work to produce a virtual meter



 5

5

Basic Ideas of the Project

- At every moment, we know which well is attached to the separator. We copy its tags to a set of global tags.
- Models are trained for the oil, gas, and water phases separately for all those points selected above – the separator output tags are the training data.
- We thus obtain three tags for each well with the three phases continuously, whether connected to the separator or not.
- For the existing well tests, we compare the computed against the measured output to judge the accuracy of our virtual flow meters.

 6

6

Example: Gas Flow

- The gas flow rate as measured, compares well with the model in this case.

© 2021 7

7

Well Tests

- When a well is connected to the separator, the oil, gas, and water flows are measured.
- The gas flow value is then **corrected** because gas is injected into the well. As the gas lift rate tag is not reliable, this correction cannot be undone by us.
- We therefore use the data directly from the separator as our baseline (**raw baseline**), and not the data found in the well test report (**reference baseline**).
- Measurement and Model are compared in percent relative to measurement.

© 2021 8

8

Example Well Test

Well 23 was tested between 2018-01-08 19:00:00 and 2018-01-09 05:00:00.

	Reference	Raw	Model	Error (%)
Oil	39.792	39.931	45.034	13
Gas	43771.2	190835.0	198124.9	-3
Water	1225.3	1224.7	1193.4	4
Water Cut	0.969	0.968	0.964	0.4
GOR	1100	4779	4399	-8

Corrected for gas injected.

Baseline for our study

© 2021 9

9

Method

algorithmica technologies © 2021 10

10

Every Well has these Tags …

1. Downhole Pressure – some wells lack this
2. Downhole Temperature – some wells lack this
3. Wellhead Pressure
4. Wellhead Temperature
5. Production Choke
6. Pressure downstream choke
7. Gaslift Rate – measurement is unreliable
8. Gaslift Pressure
9. Gaslift Choke
10. Flowline Pressure

These tags form the raw input data that we have available once per minute.

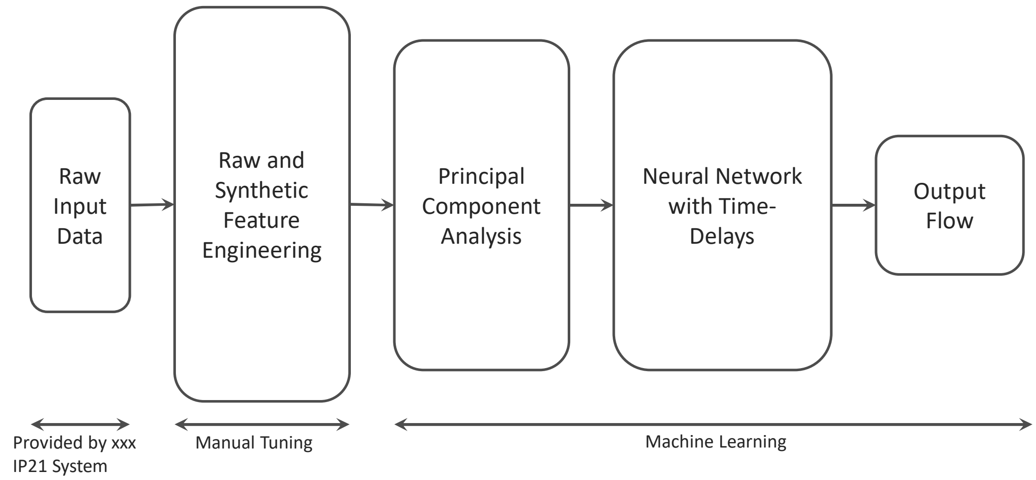
The Separator measures …

1. Oil Flow,
2. Water Flow, and
3. Gas Flow

… for the well that is currently connected to the separator every minute.

These tags form the raw teacher output data that we have used to train the models.

Overall Procedure – for each Phase separately



13

13

Major Scientific Conclusions

- The 10 raw input tags are not sufficiently informative on their own to allow an accurate machine learning to take place.
- It was necessary to construct a moving window for each raw tag of length 20 minutes and to compute the average, variance, skewness, and kurtosis for it.
- It was also necessary to introduce a time-delay between the well's tag and the known separator output. Oil has an inherent time delay of approximately 40 minutes, Gas 20 minutes, and Water 10 minutes. The causal reason is unknown. It was deduced from the correlation functions of all tags.
- With these synthetic features, suitably delayed, the data is informative. In order to prevent a too large model (and thus overfitting), the dimensionality was reduced using principal component analysis. This reduced the dimensionality by about 50% while keeping about 99% of the data variance.

14

14

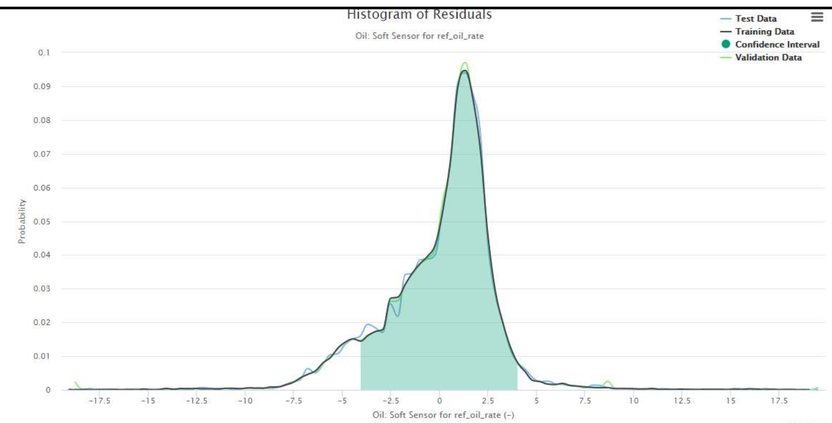
General Accuracy Measures

- A residual is the difference between the measured and the modeled value.
- We graph the frequency of the residual (vertical axis) against the value of the residual (horizontal) axis to get a distribution/histogram.
- If the histogram is a bell-shaped curve, symmetrical, with its peak approximately at zero, then the model is basically sound.
- The width of this distribution is roughly the expected accuracy of the model. Other statistical measures can be computed from the distribution.

 15

15

Residuals: Oil Flow

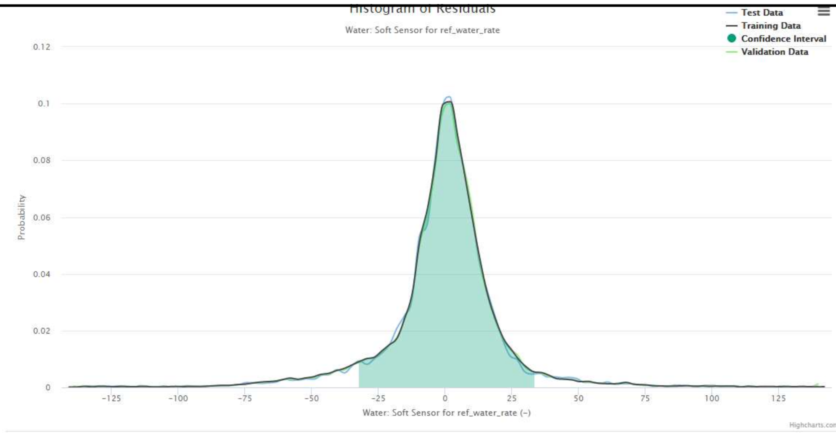


	Training Data	Test Data	Validation Data
Samples	56593	10011	67417
R^2	0.791058	0.789374	0.788375
RMSE	0.032029	0.032577	3.245099
Residual Mean	0.036079	0.052397	0.017659
Mean Absolute Residual	2.135766	2.15519	2.208776
Residual Std.Dev.	2.899805	2.938082	3.245033
Residual Skewness	-0.516724	-0.48516	-1.469175
Residual Kurtosis	4.645243	4.948554	17.796669

 16

16

Residuals: Water Flow

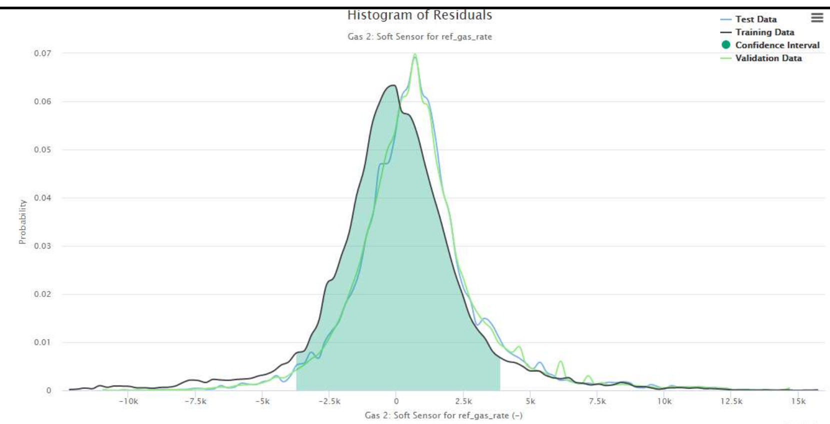


	Training Data	Test Data	Validation Data
Samples	56606	9979	67417
R^2	0.834588	0.835115	0.836245
RMSE	0.064345	0.063354	23.839907
Residual Mean	-0.104725	-0.127686	0.059638
Mean Absolute Residual	14.556548	14.333103	14.655396
Residual Std.Dev.	23.416861	23.241714	23.839865
Residual Skewness	-0.078995	-0.251926	0.054437
Residual Kurtosis	7.49489	7.772028	8.536174

 17

17

Residuals: Gas Flow



	Training Data	Test Data	Validation Data
Samples	56222	10107	67417
R^2	0.900564	0.670606	0.663429
RMSE	0.07938	0.075469	2516.182373
Residual Mean	-12.765525	742.790405	771.582947
Mean Absolute Residual	1823.569092	1735.757568	1753.012329
Residual Std.Dev.	2634.227051	2361.949707	2394.943604
Residual Skewness	0.046712	0.835122	0.945531
Residual Kurtosis	3.778615	3.793585	4.3096

 18

18

Results

algorithmica technologies © 2021 19

19

Summary of Well Tests

- Full details are available as an Excel sheet. Here are some high-level results from that table.
- There were 158 relevant well tests that were selected as the basis for this study.
- Most well tests were reproduced with good accuracy. Some well tests – when very little of one phase arrived at the separator – were reproduced poorly.
- The individual phases are calculated well. The water cut is dominated by water and is accurate too. The GOR gets deviations from gas and oil and is therefore not very accurate.

© 2021 20

20

Accuracy of Well Tests

The number of well tests that are more accurate than a certain percentage for that KPI. The "sweet spot" is circled, so that we may say oil is accurate to 15%, water to 8%, and gas to 10%. Total number of well tests was 158.

	Oil	Water	Gas	Water Cut	GOR
No. Well Tests accurate to 6%	65	112	96	155	52
No. Well Tests accurate to 7%	74	130	104	156	58
No. Well Tests accurate to 8%	85	153	119	156	78
No. Well Tests accurate to 10%	107	153	156	156	95
No. Well Tests accurate to 15%	152	153	156	156	131

Thus, 152 of 158 well tests have the oil model within 15% of the actual oil flow at the separator; 153 of 158 well tests have the water model within 8% of the actual water flow; 156 of 158 well tests have the gas model within 10% of the actual gas flow.

 21

21

Documentation of Results Table

Columns	Explanation
A, B, C, D	Specifies when the well test began and ended as well as how long it took and which well was being tested.
E, F, G	The oil, water, and gas flows as written in the well test report by Wintershall Dea that includes the gas flow correction. We call this the reference data.
H, I, J	The oil, water, and gas flows derived from the IP21 raw data coming from the separator tags that does not include the gas flow correction. We call this the raw data.
K, L, M	The oil, water, and gas flows according to the machine learning models and trained on the IP21 data.
N, O, P	The errors (in %) of the oil, water, and gas flow models relative to the raw data.
Q, R, S, T	The water cut as computed by Water/(Water + Oil) of the reference, raw, and model values, as well as the error of the model water cut relative to the raw water cut.
U, V, W, X	The gas-oil-ratio as computed by Gas/Oil of the reference, raw, and model values, as well as the error of the model gas-oil-ratio relative to the raw gas-oil-ratio.

 22

22

Conclusion

Conclusion of Proof of Concept

- Out of 158 relevant well tests, the models for oil, water, and gas flows were found to produce accurate results for all but a few tests.
- The well tests that were reproduced poorly were those where little of the relevant phase was extracted.
- The models exist for each well and could be brought online to provide continuous virtual well testing for every well – after some software development.
- Should a new well be drilled, this well could be instrumented immediately with no effort. Models can be tuned periodically in the future as more reference well tests become available.

Next Steps

- If xxx would like to proceed from this PoC to a deployment, these would be the next steps …
- Workshop with xxx to design the right graphical user interface for the multiphase flow meters.
- Implementation of the user interface, polishing of data interfaces, and testing of the full computational flow.
- User feedback sessions in an agile manner to achieve the right usability with all the required outputs.

 25

25

Hands on Project

- Work with a curated dataset from MPFM-equipped wells
  - 67,418 data-samples with the following 16 measurements
    - Date-time, well, reservoir
    - chokeprod, chokegaslift, chokepressdownstream
    - ref_oil_rate, ref_water_rate, ref_gas_rate , gasliftrate
    - dht, dhp, wht, whp, gasliftpressure, flowlinepressure
- It's a Regression problem
  - Perform Data exploration, correction-interpolation (as needed)
  - Develop regression models with various ML-methods

26

Pan American ENERGY

14.0. Closing Remarks…

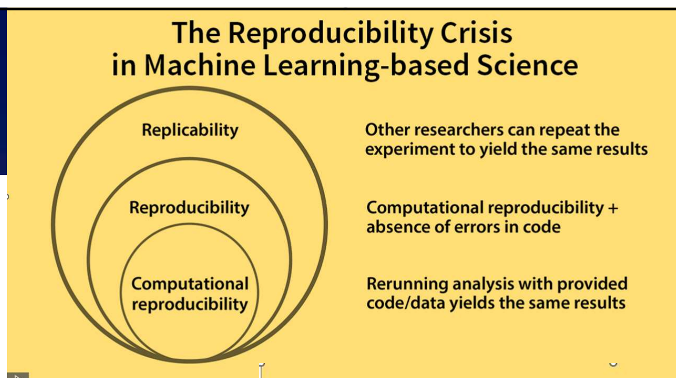
Dr. Rajan Chokshi

Data Analytics for Artificial Lift & Production Optimization

BAUERBERG KLEIN E&P TRAINING SOLUTIONS

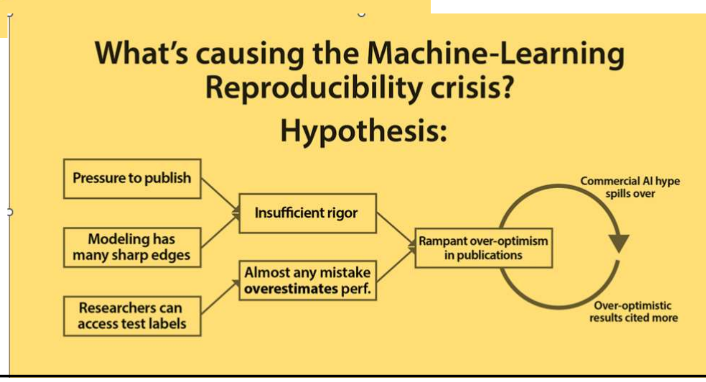
Accutant Solutions
Accurate Accountable Acumen

1



Science-plagued-by-machine-learning-mistakes-deep fakes-censor-profanity-….

Source: The Batch, Aug 17 2022, DeepLearning.AI, https://bit.ly/3Ca0Eot Accessed Aug 17 '22



2

Avoid ML Pitfalls – Some Dos and Don'ts for academics and **all** beginners

2 Before you start to build models
- 2.1 Do take the time to understand your data
- 2.2 Don't look at *all* your data
- 2.3 Do make sure you have enough data
- 2.4 Do talk to domain experts
- 2.5 Do survey the literature
- 2.6 Do think about how your model will be deployed

3 How to reliably build models
- 3.1 Don't allow test data to leak into the training process
- 3.2 Do try out a range of different models
- 3.3 Don't use inappropriate models
- 3.4 Do optimise your model's hyperparameters
- 3.5 Do be careful where you optimise hyperparameters and select features

4 How to robustly evaluate models
- 4.1 Do use an appropriate test set
- 4.2 Do use a validation set
- 4.3 Do evaluate a model multiple times

4 How to robustly evaluate models
- 4.4 Do save some data to evaluate your final model instance
- 4.5 Don't use accuracy with imbalanced data sets

5 How to compare models fairly
- 5.1 Don't assume a bigger number means a better model
- 5.2 Do use statistical tests when comparing models
- 5.3 Do correct for multiple comparisons
- 5.4 Don't always believe results from community benchmarks
- 5.5 Do consider combinations of models

6 How to report your results
- 6.1 Do be transparent
- 6.2 Do report performance in multiple ways
- 6.3 Don't generalise beyond the data
- 6.4 Do be careful when reporting statistical significance
- 6.5 Do look at your models

Source: Lones, M., "How to avoid machine learning pitfalls: a guide for academic researchers", https://arxiv.org/pdf/2108.02497.pdf

3

What should be our approach?

Source: Andrew Ng at Amazon re:MARS 2019, https://youtu.be/j2nGxw8sKYU?t=615

4

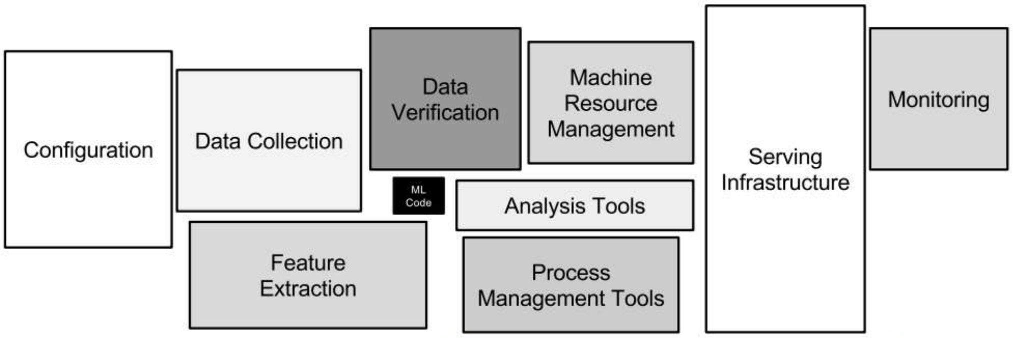


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Hidden Technical Debt in Machine Learning Systems

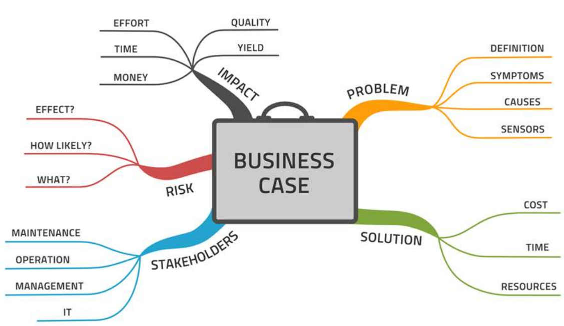
D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.

Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison
{ebner,vchaudhary,mwyoung,jfcrespo,dennison}@google.com
Google, Inc.

Source: Sculley et al, "Hidden Technical Debt in Machine Learning Systems", https://bit.ly/3Avp76u accessed Aug 17 22

5

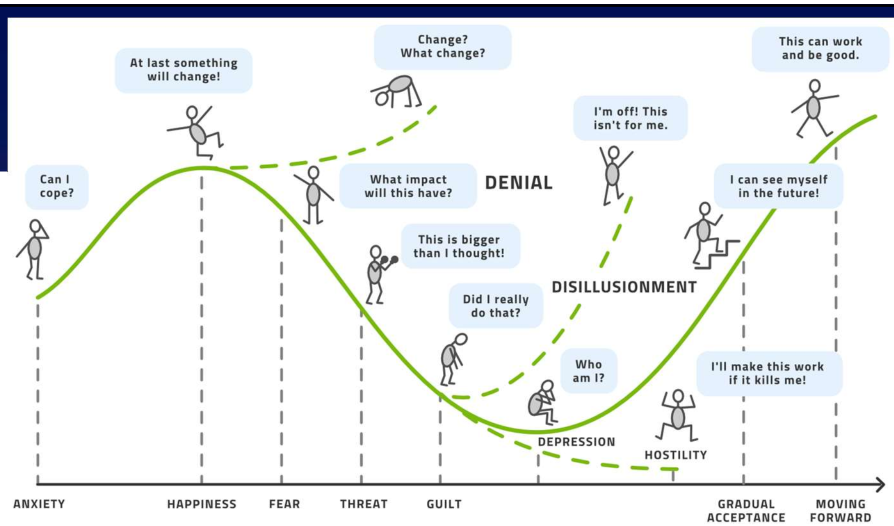
Do first things first….
Formulate Business Case, Ask Questions



1. If we had a model, what impact would it have?
2. How accurate does it need to be for that impact?
3. Who needs to be convinced for it to be used?
4. What are the risks?
5. What are the costs and resource needs?

Courtesy: Algorithmica

6

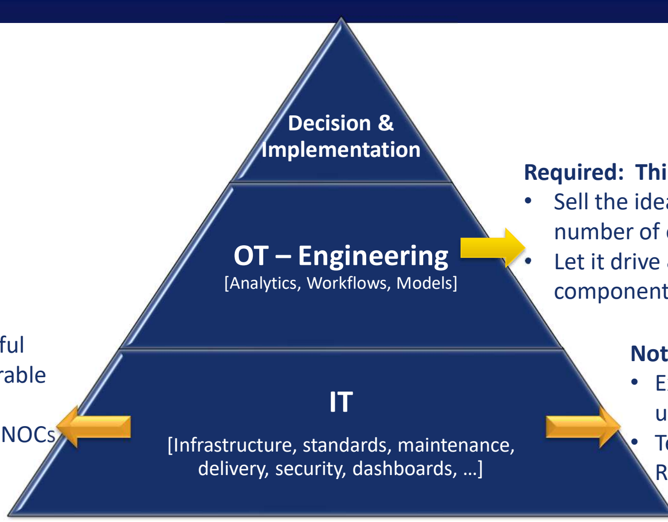


Machine Learning is not an end onto itself. It lives in an ecosystem of human factors and must be embedded in order to deliver value. Management and Change-Management is essential.

Courtesy: Algorithmica

7

DOF Stakeholders : IT & OT



Required: This is the prize
- Sell the idea & technology to large number of engineers, scientists
- Let it drive & determine the scope of IT component

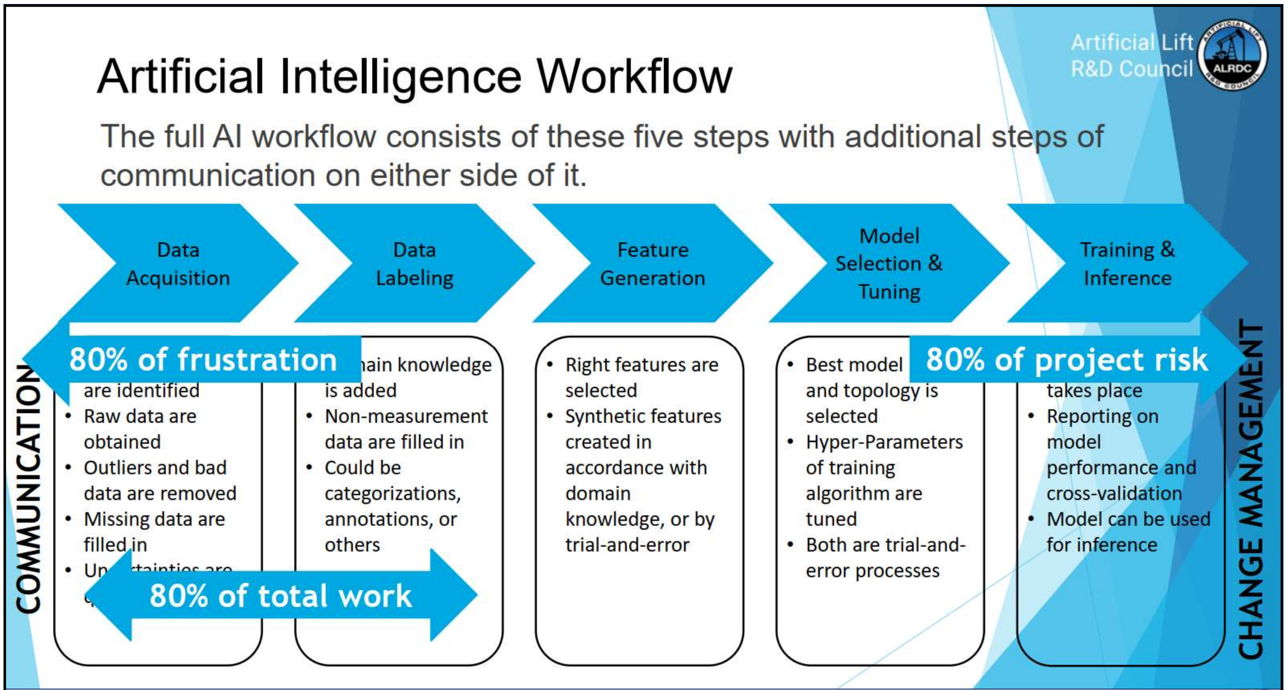
Required, once
- Technology-test: Successful
- Management view: Favorable
- Resistance: Minimal
- Streamline must: IOCs & NOCs

Not Required, while
- Experimenting & assessing the utility
- Technology is not completely sold
Resistance high

Adapted from SPE webinar: Transforming E&P Applications through Big Data Analytics, Dr. Mohaghegh, http://bit.ly/1O5rVEw

8

8



Source: Keynote address "The Status and Workflow of Artificial Intelligence for Artificial Lift," Patrick Bangert, Samsung SDS, ALRDC 2021 International SRP Workshop, Feb 8-12 21

9

Conclusion

- Data driven approaches are increasingly being applied in the artificial lift domain
- Prevalent applications cover anomaly detection, failure prediction and virtual flow metering
- The use of ML helps save time, minimize effort, improve quality, increase yield, limit human error, increase accuracy and lower risk.

10

Takeaways

- While ML/AI approaches promise new pathways to solving our operational and design challenges, AL and the production community needs to actively pay attention to and participate in the data lifecycle from cleaning through modeling to production-deployment and retraining.

11

Contact Details

SCAN ME

- Dr. Rajan Chokshi

  rajan.chokshi@accutants.com

  https://Linkedin.com/in/RajanChokshi

12

Feedback Survey from SPE

13

Dr. Rajan Chokshi boasts a career with four decades of artificial lift, real-time production optimization, software development, and management experience. He is currently engaged in optimizing artificial lift, calculating multi-phase flow, employing ML/AI for failure prediction, designing virtual flow meter, and managing competency. Chokshi has gained experience from national oil companies, majors, independents, service providers and technology companies. He has collaborated on over fifteen SPE papers, gotten two patents, and presented multiple SPE webinars and several training courses, including graduate courses at universities.
He has been selected twice as an SPE distinguished lecturer and is also actively taking part in several SPE technical committees.Dr. Chokshi has attained a Bachelor's and Master's Degree in chemical engineering from Gujarat University and IIT-Kanpur, India, and a Ph.D. in Petroleum Engineering from the University of Tulsa, USA.

Worldwide Experience in Training / Seminar / Workshop Deliveries:

- Several SPE webinars, ALRDC and SPE trainings globally on artificial lift and production optimization themes.
- Graduate courses at Texas Tech, Missouri S&T, Louisiana State, PDPU, U of Southern California, and U of Houston.
- Bespoke trainings / workshops and seminars globally for practicing professionals.
  - **Companies:** Apache, Basra Oil Company, Chevron, Cimarex, ConocoPhillips, EcoPetrol, Equinor, KOC, ONGC, LukOil, Newfield Exploration, Occidental Petroleum, PDO, PDVSA, PEMEX, PetroCanada, Petronas, Qatar Petroleum, Repsol, Saudi Aramco, Shell, Sonatrech, Tatneft, United Energy Pakistan, Vista Energy, YPF, and others.
  - **Countries:** USA, Algeria, Argentina, Bahrain, Brazil, Canada, China, Croatia, Congo, Ghana, India, Indonesia, Iraq, Kazakhstan, Kenya, Kuwait, Libya, Malaysia, Mexico, Oman, Norway, Qatar, Romania, Russia, Saudi Arabia, Serbia, Singapore, S Korea, Tanzania, Thailand, Tunisia, Turkmenistan, UAE, Ukraine, Uzbekistan, Venezuela.
  - Virtual training provided to global audiences since pandemic.

List of Courses

Nodal Analysis for Oil & Gas Wells
Artificial Lift and Production Optimization
Artificial Lift and Optimization for Unconventional Assets
Artificial Lift and Real-Time Optimization in Digital Oilfield

Gas Lift Design & Optimization using NODAL Analysis
Gas Lift Optimization in Unconventional
Gas-Lift & Deliquification Applications

Advanced Sucker Rod Pumping
Advanced Sucker Rod Pumping – Unconventional Wells

Electrical Submersible Pumping using NODAL Analysis
Advanced Artificial Lifting with ESP
Optimization of ESP & Gas-Lift
Optimization of ESP & Sucker Rod Pumping

Gas Well Deliquification and Production Optimization

Data Analytics Workflows for Artificial Lift, Production and Facility Engineers