Risk Management in the Energy Industry

Susan Nash, Ph.D.

Unit 3: Predicting Risk with Artificial Intelligence and Machine Learning

Unit 3: Predicting Risk with Artificial Intelligence and Machine Learning

Unit Learning Objectives

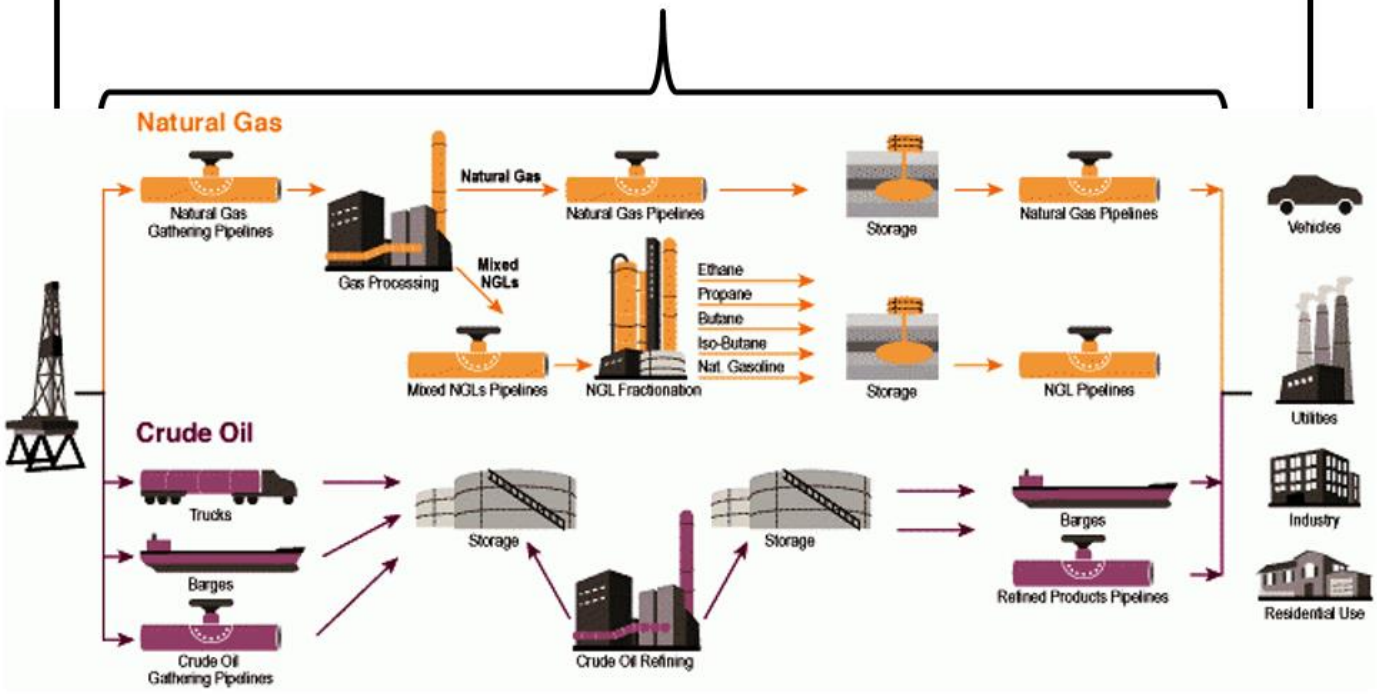
- Upon successful completion of this unit, learners will be able to identify how to use artificial intelligence and machine learning to predict levels and types of risk, both known and unknown.

ML in the Energy Industry

- **Upstream**
- Classify wells using your own unique set of criteria
- Identify high-value (or potential high-value) blocks
- **Midstream**
- Classify infrastructure (pipelines, etc) with your own criteria
- Predict performance and bottlenecks
- **Downstream**
- Bottlenecks / Refining / Retail and distribution risk

ML in the Petroleum Industry

- Upstream
  - Drilling
  - Production
- Midstream
  - Transportation
  - Treatment
- Downstream
  - Industrial
  - Residential
  - Commercial



Workflow for Machine Learning

- Pinpoint the problem you want to solve.
- Identify the data you'll need to use
- Collect the data
- Clean the data
- Organize your data (put into a model - if structured, may use Open Source models such as those from Apache Hadoop)
- Find a model
- Develop algorithms (May use repositories and also cloud-based interfaces)
- Train the model
- Test with data sets
- Reality check
- Decision points

What is "Dirty" Data

- Does not make sense
- Bad labels
- Incorrect formatting
- Too many "nulls"
- Missing information

	DBAName	AKAName	Address	City	State	Zip
t1	John Veliotis Sr.	Johnnyo's	3465 S Morgan ST	*Chicago*	IL	*60608*
t2	John Veliotis Sr.	Johnnyo's	3465 S Morgan ST	Chicago	IL	*60609*
t3	John Veliotis Sr.	Johnnyo's	3465 S Morgan ST	Chicago	IL	*60609*
t4	*Johnnyo's*	Johnnyo's	3465 S Morgan ST	*Cicago*	IL	60608

Conflicts

Does not obey data distribution

Conflict

Oldest Book : Select Query

Author	TITLE	YEAR
Ed: William B. Yeats	Irish Folk Stories & Fairy Tal	0
Hawthorne, Nathaniel	Twice-Told Tales	0
Ransome, Arthur	Russian Fairy Tales	0
Davis, Washington	Campfire Chats Of Th Civil War	1884
Andersen, Hans Chris	Andersen's Marchen	1893
Storm, Theodor	Immensee	1902
Stratton-Porter, Gene	Freckles	1904
Alarcon, Pedro	Novelas Cortas Escogidas	1905
Stratton-Porter, Gene	Harvester, The	1912
Gerstacter, Friedrich	Germelshausen	1913

Use Excel or Python to Clean Data?

- Use Excel if:
- You have fewer than 1 million records
- You need to do the job in a way that is quick and easy
- There is a logical pattern to cleaning the data and it's easy enough to clean using Excel functions
- The logical pattern to cleaning the data is hard to define, and you need to clean the data manually

Use Excel or Python to Clean Data?

- Use Python if:
- You need to document your process
- You plan on doing the job on a repeated basis
- There is a logical pattern to cleaning the data, but it is hard to implement with Excel functions

- Brendon Bailey. "Data Cleaning 101" TowardDataScience.com https://towardsdatascience.com/data-cleaning-101-948d22a92e4

Data Storage

- *There are a number of Cloud options*
- Google Cloud
- Amazon Web Services (AWS)
- Dell EMC
- Microsoft Azure

Cloud Environments & Tools for Risk Analysis

- Spotfire (http://www.spotfire.com)
- Qlik.com (free Spotfire alternative, Qlik.com)
- Jupyter Notebook https://jupyter.org/
- iPython, R, C++, Julia, etc.
- A Gallery of interesting Jupyter Notebooks (ready to share) https://github.com/jupyter/jupyter/wiki/A-gallery-of-interesting-Jupyter-Notebooks

Predict Where & When High-Risk Events Might Take Place

- Define the scope
- Gather data
- Clean data
- Put data in structure
- Use probabilistic analysis (Spotfire, etc.)
- Include geospatial

Introducing TensorFlow

- TensorFlow™ is an open source software library for high performance numerical computation. Its flexible architecture allows easy deployment of computation across a variety of platforms (CPUs, GPUs, TPUs), and from desktops to clusters of servers to mobile and edge devices.

- Originally developed by researchers and engineers from the Google Brain team within Google's AI organization, it comes with strong support for machine learning and deep learning and the flexible numerical computation core is used across many other scientific domains.
- https://www.tensorflow.org/
- https://www.tensorflow.org/install/

Learn & Use Machine Learning with TensorFlow

- Tensorflow: https://www.tensorflow.org/tutorials/keras/
- Tensorflow Machine Learning Cookbook: https://github.com/nfmcclure/tensorflow_cookbook
- AI and Probabilistic Models
- Part I
- https://medium.com/tensorflow/industrial-ai-bhges-physics-based-probabilistic-deep-learning-using-tensorflow-probability-5f215c791863
- Part II
- https://medium.com/tensorflow/predicting-known-unknowns-with-tensorflow-probability-industrial-ai-part-2-2fbd3522ebda

Convolutional Neural Networks to Solve Mineral Prospectivity

- Framing the exploration task as a supervised learning problem, the geological, geochemical and geophysical information can be used as training data, and known mineral occurrences can be used as training labels. The goal is to parameterize the complex relationships between the data and the labels such that mineral potential can be estimated in under-explored regions using available geoscience data.

- Granek, Justin. (2016). Application of Machine Learning Algorithms to Mineral Prospectivity Mapping. Open Collections. University of British Columbia. https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0340340

Machine Learning Applications: Excellent Resources

- Bougher, Benjamin Bryan. (2016) Machine Learning Applications to Geophysical Data Analysis. Open Collections. University of British Columbia. https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0308786

- Bougher, Ben B. (2016) Using the scattering transform to predict stratigraphic units from well logs. Seismic Laboratory for Imaging and Modeling (SLIM), The University of British Columbia, Vancouver https://www.slim.eos.ubc.ca/Publications/Public/Journals/CSEGRecorder/2016/bougher2015CSEGust/bougher2015CSEGust.html

- Data: Trenton Black River gamma ray logs
  Methodology: supervised learning ("uses labelled datasets to train a classifier to make predictions about future data" (Bougher, 2016))

- Methodology - what's the algorithm? Bougher uses a scattering transform - and then it feeds a K-Nearest Neighbours (KNN) classifier).

Sources of Free Petroleum Exploration and Production Data

- A report with links to free, open source, and government data used in petroleum exploration
- https://www.hgs.org/node/5923
- A website with links to public data:
- https://websites.pmc.ucsc.edu/~brodsky/wellindex.html
- Example: Kansas Geological Survey: http://www.kgs.ku.edu/PRS/petroDB.html