The Quality 4.0 Revolution:
Reveal Hidden Insights Now With Data Science and Machine Learning

Nicole M. Radziwill, PhD, MBA
ASQ Fellow & Editor, *Software Quality Professional*
Quality Practice Leader

INTELEX
Cameras on front of remote-controlled cars sampled at 20 images/second

Based on patterns in the photos, give car steering instructions

**Goal**: Keep them on the road up to max speed!

(AWS with multilayer neural net in Python/TensorFlow in Jupyter notebooks)

“... a leading car manufacturer... was experiencing recurrent faults, and therefore high warranty costs, on the rear tail-lights. The engineers had been looking at the rear of the vehicle for the answer (and not succeeding) however the software quickly found the root cause of the fault to be located in the roof of the vehicle, a part of the car that had not even been investigated as a possible source. Sometimes prior experience, or being too close to a problem, can inhibit a solution if an old hypothesis is applied to a new problem.”

From http://www.qmtmag.com/display_eds.cfm?edno=9665074
Figure 5

Burden Top IR Image
(real time, live)

Average Top Temp

Temp. Profile along the white line
(virtual above burden probe)

Figure 7

Size of the central gas flow: large chimney (a), medium chimney (b), small chimney (c), no clear chimney (d).

Figure 8

Shape of the central gas flow: solid shape (a), hollow or donut shape (b), “G” shape (c), “G” shape (d).

“chimney”

You will:

1. Understand **what Quality 4.0 is** and the value propositions it presents.

2. Identify the **types of problems** to which you can apply these new methods in your organization.

3. Explain how these technologies support **engagement and collaboration** to promote a connected and empowered workforce.

4. Explore **ways to incorporate data science and machine learning into your work**, whether you are drowning in data or just starting your digital transformation.
concerned parent: if all your friends jumped off a bridge would you follow them?
machine learning algorithm: yes.

12:20 PM - 15 Mar 2018

2,540 Retweets  5,045 Likes

Shane Celis @shanecevis  13h
Replies to @computerfact
machine learning algorithm: As a bridge-jumping enthusiast, have you considered jumping off these other fine bridges?

1  15  87
Nicole Radziwill

Quality Practice Lead, Intelex
Fellow, American Society for Quality (ASQ)
CSSBB #11962 & CMQ/OE #9583
Ph.D. Quality Systems, Indiana State
Editor, Software Quality Professional

Previously:
• Division Head, Software, Green Bank Observatory, 2002-2006
• Assistant Director of End to End Operations, National Radio Astronomy Observatory, 2006-2009
• Associate Professor of Data Science & Production Systems, James Madison University, 2009-2018; taught data science & applied machine learning from 2009-2018
1: Quality 4.0
and value propositions enabled by data science
Quality as:
1.0 - Inspection
2.0 - Design
3.0 - Engagement
<table>
<thead>
<tr>
<th>Quality Planning</th>
<th>Quality Assurance/ Quality Control</th>
<th>Quality Improvement</th>
</tr>
</thead>
</table>
| • Establish quality goals  
• Understand organizational context (mission, vision, values, strategy, position, capabilities)  
• Understand business environment  
• Identify customer & stakeholder needs – Voice of Customer (VoC)  
• Design products/services to meet needs, implement design controls  
• Design processes for consistent production & related controls | • Evaluate process performance  
• Compare performance with quality goals  
  • Manage “quality events” and update controls  
  • Audits  
  • Management reviews  
  • Nonconformances  
  • Incidents/near misses  
  • Complaints  
• Out-of-control production  
• Act to close the gaps | • Build capabilities (people and technical)  
• Build infrastructure  
• Identify and justify needs and gaps  
• Conduct improvement activities and establish controls to maintain them  
  • Quick wins (PDCA)  
  • Process improvement (DMAIC, Lean)  
  • Process design or redesign (DMADV)  
  • Business process mgmt (BPM) or Robotic Process Automation (RPA) |
Quality Events indicate that quality goals are not being met and action is needed.

- Nonconforming product
- Incidents/near misses
- Customer complaints
- Audit findings
- Recalls/warranty calls
- Deviations (from SOP)
- Out-of-control Action Plans
- Industry-specific events (e.g. MDRs)

Flowchart:

- Serious or systematic?
  - yes: CAPA
  - not really: containment
Quality Controls to prevent or correct unwanted or unexpected change → stability and consistency

- Calibrations
- Maintenance
- Inspections
- Sampling incoming parts
- Process validation
- Mistake-proofing
- In-situ process monitoring
- Environment monitoring
- Professional testing/competency assessment
- Training programs and reminders
- Corrective actions taken
- Information security/network security
Why Quality Systems Work

## AI/ML Can Enhance & Accelerate Organizational Learning

<table>
<thead>
<tr>
<th>People</th>
<th>Intelligent Systems</th>
</tr>
</thead>
</table>
| As people learn about the work environment, they update:  
  • policies,  
  • procedures,  
  • practices, and  
  • heuristics  
  to improve the performance of people, projects, processes & the bottom line. | When AI/ML systems learn, they update how they:  
  • predict,  
  • classify,  
  • find patterns, and  
  • identify important predictors  
  to improve the performance of people, projects, processes & the bottom line. |

* Jim Duarte (ASQ TV) also noticed that the algorithms do what we’ve been used to doing as quality professionals… just iteratively, and much faster.*
Quality as:

1.0 - Inspection
2.0 - Design
3.0 - Engagement
4.0 - Discovery

AI/ML evolves data-driven decision making to be self-aware & adapt to changing environments, circumstances, and customer/stakeholder needs
Data Science:
- Predict
- Identify Causal and Noncausal Relationships
- Drive Value

Through:
- Data Aggregation
- Real-time pipelines
- Dynamic Modeling

→ Learn
Hard Drive Cost per Gigabyte
1980 - 2009

$30/GB

$~0/GB

From http://www.mkomo.com/cost-per-gigabyte
c/o 92!

**Computer Science**
- Big Data infrastructure
- Programming/data access
- Deploying ML models
- Streaming data

**Math & Statistics**
- Machine learning (ML)
- Model building and evaluation
- Anomaly detection
- Forecasting
- Ensembles
- Design of Experiments

**Domain Expertise**
Scientific Method vs Data Science

1. Develop a hypothesis based on theory
2. Take a random, representative sample from the population
3. Compare characteristics to your prediction
4. Draw conclusions

Traditional

1. Just look at the entire population of data and see what you find… we have the tools to do it and the data is cheap! Who needs theory or a hypothesis?

Data Science

‘…GFT overlooks information that could be extracted by traditional statistical methods.’

‘“Big data hubris” is the often implicit assumption that big data are a substitute for, rather than a supplement to, traditional data collection and analysis.’

What is Big Data?

• Lots of it (**volume**)
• It’s coming at you fast (**velocity**)
• Different formats and sampling frequencies (**variety**)
• Huge variations in data quality (**veracity**)
• Different people/organizations produce it or own it (**governance**)
• It could easily change or disappear (**control**)
• There may be restrictions on how you use it (**policy**)

“… Big Data is anything bigger or more complex than what your organization is currently prepared to handle.”

-- one of the world’s experts on cyberinfrastructure for Big Data at a National Science Foundation panel (2013)
<table>
<thead>
<tr>
<th>Type of data</th>
<th>Inventory management</th>
<th>Transportation management</th>
<th>Customer and supplier relationship management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>How can sales data be used with detailed customer data to improve inventory management either in terms of forecasting or treating some inventory as “committed” based on specific shoppers requirements?</td>
<td>How can more current sales data be used to re-direct shipments in transit? How can sales data, integrated with detailed customer data, be used for more efficient and effective merge-in-transit operations?</td>
<td>How can more granular sales data from the wide variety of sources that exist be used to improve visibility on the one hand and trust on the other, between trading partners?</td>
</tr>
<tr>
<td>Consumer</td>
<td>How can face profiling data for shopper identification, emotion detection, and eye-tracking data be used to determine which items to carry and stock at particular shelf locations?</td>
<td>How can delivery preferences captured in online purchases be used to manage transportation mode and carrier selection decisions?</td>
<td>How can customer sentiment about products purchased based on “Likes,” “Tweets,” and product reviews be used to collaborate on forecasts?</td>
</tr>
<tr>
<td>Location and time</td>
<td>How can sensor data used to detect location in store be used to improve inventory management, including departmental merchandising decisions?</td>
<td>How can sensor data in the distribution center be used to anticipate transportation requirements?</td>
<td>How can location and time-stamp data of shoppers be used for collaborative assortment and merchandising decisions?</td>
</tr>
</tbody>
</table>

Degrees of Automation

1. Human specifies process and computer only implements
2. Computer assists by determining options
3. Computer assists by determining options and suggests choice
4. Computer assists by determining options and selects choice - human has option to do it, or not
5. Computer selects and implements options if human approves
6. Computer selects option and automatically implements, but gives human chance to stop the process flow
7. Computer selects and implements options, then tells human about results
8. Computer selects and implements options, but only tells human about results if asked for report
9. Computer selects and implements options, but only tells human about results if asked, only reports some information
10. Computer performs whole job; may or may not tell human

2: Types of Problems
& how to identify challenges that are well supported by DS/ML
Sit down if you think your org has solved any of these problems for front line staff. Narrator: only 2 people out of 200 sat down. #quality2018

**What problems are we trying to solve?**

What challenges do frontline leaders in health care face?

- Time pressed - high service demand & patient complexity
- Data rich but information poor
- Unclear on what to prioritise
- Communication challenges
- Variability in quality of care & patient experience
- Inconsistent processes & inefficiencies
- Staff time not always well utilised
- Resource constrained

.....Putting out fires all day
Quality 4.0 Ecosystem

Machine Learning Tasks

1. Prediction
2. Classification
3. Pattern Identification
4. Data Reduction
5. Anomaly Detection
6. Pathfinding

Data Science Tasks

1. Frame the problem(s)
2. Explore the data
3. Build models
4. Evaluate model performance
5. Deploy models and real-time processing **pipelines**
6. Uncover high-value insights
7. Tell stories
8. Do it again… and again
“Data can take several forms, such as data at rest and data in motion. Data that are sitting in warehouses waiting to be analyzed are referred to as data at rest…

**Data in motion** (data analyzed in real time as the event occurs, such as click streams and sensors) flow through a connected device to a database on the receiving end for immediate analysis… used to fine tune the [predictive] models… and analytics at the edge… deploys information to control processes in real time.”

-- Jim Duarte

<table>
<thead>
<tr>
<th>Quality Problem</th>
<th>Machine Learning</th>
<th>Feasibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio anomalies</td>
<td>Classification (CNN)</td>
<td>Feasible</td>
</tr>
<tr>
<td>Checking for leakage</td>
<td>Image classification</td>
<td>Potential</td>
</tr>
<tr>
<td>Setting up engines</td>
<td>Reinforcement learning</td>
<td>Unnecessarily</td>
</tr>
<tr>
<td>Supervising testing</td>
<td>Image classification</td>
<td>complicated</td>
</tr>
<tr>
<td>Feature detection</td>
<td>Image classification</td>
<td>Unnecessarily</td>
</tr>
<tr>
<td></td>
<td>Multidimensional</td>
<td>complicated</td>
</tr>
<tr>
<td>Cause diagnosis</td>
<td>Regression, association</td>
<td>Potential</td>
</tr>
<tr>
<td></td>
<td>rule learning</td>
<td></td>
</tr>
<tr>
<td>Machine adjustments</td>
<td>Multidimensional</td>
<td>Potential</td>
</tr>
<tr>
<td></td>
<td>Regression</td>
<td></td>
</tr>
<tr>
<td>Surface defects</td>
<td>Image classification</td>
<td>Feasible</td>
</tr>
<tr>
<td>Assembly gaps</td>
<td>Image Regression</td>
<td>Unnecessarily</td>
</tr>
<tr>
<td></td>
<td></td>
<td>complicated</td>
</tr>
<tr>
<td>Part identification</td>
<td>Image classification</td>
<td>Unnecessarily</td>
</tr>
<tr>
<td></td>
<td></td>
<td>complicated</td>
</tr>
<tr>
<td>Machine error cause</td>
<td>Image classification</td>
<td>Unnecessarily</td>
</tr>
<tr>
<td></td>
<td></td>
<td>complicated</td>
</tr>
<tr>
<td>Field quality detection</td>
<td>Association rule</td>
<td>Unnecessarily</td>
</tr>
<tr>
<td></td>
<td>learning</td>
<td>complicated</td>
</tr>
<tr>
<td>Field quality cause analysis</td>
<td>Association rule</td>
<td>Feasible</td>
</tr>
<tr>
<td></td>
<td>learning</td>
<td></td>
</tr>
<tr>
<td>Predictive maintenance</td>
<td>Classification</td>
<td>Feasible</td>
</tr>
<tr>
<td>Control chart pattern</td>
<td>Classification</td>
<td>Feasible</td>
</tr>
<tr>
<td>detection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimizing production</td>
<td>Classification</td>
<td>Potential</td>
</tr>
<tr>
<td>parameters</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Value Propositions for Quality 4.0

1. **Augment** (or improve upon) human intelligence
2. **Increase** speed and enhance the quality of decision-making
3. **Anticipate** changes, reveal biases, and adapt to new circumstances and data sources
4. **Learn** how to learn: cultivate self-awareness and other-awareness
5. **Reveal** opportunities for continuous improvement
6. **Improve** transparency, traceability, and auditability

**MACHINE LEARNING:**
Any place where **HEURISTICS** are used for decision making

**BLOCKCHAIN:**
Any place there are **TRANSACTIONS** (pre, during, post)

3: Engagement & Collaboration

with Quality 4.0

INTELEX
NOTHING CAN DEFEAT
A DIVERSE AND ENERGIZED GROUP OF PEOPLE
ENGAGED IN A SHARED VISION
AND COMMON MISSION
What is Engagement?
A Review of ~120 Research Papers

Not Just for Manufacturing, Not Just for Germany

While the focus of Plattform Industrie 4.0 has, without doubt, been very much on German manufacturing, it is by no means an exclusive club for any particular industry or region. Gartner’s recent definition of the term Industrie 4.0 underlines this point, calling it “a business-outcome-driven digital transformation approach to generate value from the collaboration of multiple partners in ecosystems across value chains and industries.”

In other words, it is not simply about technologies such as the Internet of Things (IoT) and artificial intelligence (AI) increasing production efficiency at any one site or even one company. Industrie 4.0 enables integration across entire enterprises, powering the creation of brand new ecosystems that can span multiple industries.

I have spoken many times about the importance of openness – in terms of technologies, development mindsets, partnerships, and collaborations. Industrie 4.0 is for me another great example of the crucial role openness plays in successful digitalization. On the one hand, there are the individual interorganizational networks that Industrie 4.0 helps to create – ecosystems that connect companies and industries in completely new value chains. But on a broader scale, cooperation and openness between many different types of organizations across Germany, Europe, and indeed the globe are required to fully exploit the advantages of digitalization. And among the various goals the Plattform Industrie 4.0 pursues, the promotion of networking between global consortia and stakeholders, and the coordination of this exchange is a clear priority.

4: Reveal Hidden Insights Now
incorporate DS/ML & ensure readiness for the future

INTELEX
I want my QMS/IMS to…

- Help me **make better decisions** about my processes
- **Audit itself and alert me** when I need to do something
- Tell me where to **focus** my resources for maximum added value
- Based on risk assessments and theory of constraints, tell me how should I **prioritize** work
- Alert me when an important **change** occurs at the systems level
DS in the Cloud: [http://rstudio.cloud](http://rstudio.cloud)

Example to follow: [https://qualityandinnovation.com/2015/07/14/a-simple-intro-to-bayesian-change-point-analysis/](https://qualityandinnovation.com/2015/07/14/a-simple-intro-to-bayesian-change-point-analysis/)
R version 3.5.0 (2018-04-23) -- "Joy in Playing"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

>
R version 3.5.0 (2018-04-23) -- "Joy in Playing"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(boot)
> data(coal)
> y <- tabulate(floor(coal[[1]]) [1852:1981])
> barplot(y, xlab="years", ylab="frequency of disasters")
> install.packages("changepoint")
Example 2: Wine Quality

```r
library(tidyverse)
library(randomForest)

wine <- as.tibble(read.delim(url,header=TRUE,sep=";"))

> wine
# A tibble: 4,898 x 12
   fixed.acidity volatile.acidity citric.acid residual.sugar chlorides free.sulfur.dioxide
       <dbl>             <dbl>       <dbl>             <dbl>       <dbl>            <dbl>
1         17.0            0.270        0.360            20.7         0.0450            45
2         2.0            6.30         0.300            0.340         1.60             41
# ... with 4,888 more rows, and 5 more variables: density <dbl>, pH <dbl>, sulphates <dbl>,
alcohol <dbl>, quality <int>

> wine$taste <- as.factor(ifelse(wine$quality <= 6,'bad','good'))
```

Build the ML Model

```r
> train <- wine[1:3000,]
> test  <- wine[3001:4898,]
> ml.model <- randomForest(taste ~ . - quality, data=train)
> print(ml.model)

Call:
  randomForest(formula = taste ~ . - quality, data = train)
  Type of random forest: classification
  Number of trees: 500
No. of variables tried at each split: 3

OOB estimate of error rate: 9.7%
Confusion matrix:
  bad  good class.error
bad   2276   83  0.0351844
good  208  433  0.3244930
```
Check Performance of ML Model

\[ \text{pred} \leftarrow \text{predict(ml.model, newdata=test)} \]

\[
\begin{array}{ccc}
\text{pred} & \text{bad} & \text{good} \\
\text{bad} & 1233 & 191 \\
\text{good} & 246 & 228 \\
\end{array}
\]

\[ \text{ml.model}\$importance \]

- fixed.acidity: 64.47684
- volatile.acidity: 84.14794
- citric.acid: 61.37304
- residual.sugar: 91.95227
- chlorides: 86.69503
- free.sulfur.dioxide: 82.76855
- total.sulfur.dioxide: 80.33434
- density: 125.57005
- pH: 81.95264
- sulphates: 78.01374
- alcohol: 169.45978

\[ \text{plot(ml.model, log="y")} \]

\[ \text{legend("topright", colnames(ml.model}\$\text{err.rate), col=1:4, cex=0.8, fill=1:4)} \]
Drowning in Data

Set up data **governance** processes. Bust silos, encourage collaboration, continue systems integration.

Invest in **cybersecurity**: physical, behavioral, and technical

**Be selective**: archive only most critical information, or “compute at the edges” & store aggregated values

Put **controls** in place to avoid the Data Swamp/Graveyard problem

Haven’t Started Yet

**Digitize!** Make data capture part of what’s expected in each work process

Practice good data hygiene: **one observation per row, one variable per column**

**Keep learning** key concepts, get your feet wet and play

**Wait it out!** Gartner says ~40% of data science tasks will be automated by 2020.
In a perfect world, my QMS would tell me…

INTELEX
Integrated Management Systems for Environment, Health & Safety, and Quality Management (EHSQ)


Worldwide client base.

More than 500 employees.

Over 1,200 global customers.

Over 1.5 million users.

Peer reviewed as best managed company.
Quality 4.0 is…

Connectedness

Intelligence

Automation

… for discovering insights into performance.

Nicole Radziwill
Quality Practice Lead
nicole.radziwill@intelex.com
@nicoleradziwill

https://www.linkedin.com/in/nicoleradziwill