BUILDING CAPACITY

Darryl Booth, MBA

Experimenting With Artificial Intelligence to Build Capacity

Editor's Note: **A need exists within environmental health agencies to increase their capacity to perform in an environment of diminishing resources. With limited resources and increasing demands, we need to seek new approaches to the business of environmental health. Acutely aware of these challenges, NEHA has initiated a partnership with Accela called Building Capacity—a joint effort to educate, reinforce, and build upon successes within the profession using technology to improve efficiency and extend the impact of environmental health agencies.**

The *Journal* is pleased to publish this column from Accela that will provide readers with insight into the Building Capacity initiative, as well as be a conduit for fostering the capacity building of environmental health agencies across the country. The conclusions of this column are those of the author(s) and do not necessarily represent the views of NEHA.

Darryl Booth is the general manager of environmental health at Accela and has been monitoring regulatory and data tracking needs of agencies across the U.S. for almost 20 years. He serves as technical advisor to NEHA's informatics and technology section.

Google Maps rapidly recommends the fastest route, considering vast amounts of crowd-sourced traffic data. Facebook automatically suggests the names of friends in uploaded photos and proposes you "tag" them, thereby validating its assumptions and improving future results. Mobile phones and personal voice assistants rely on voice-to-text, filtering through background noises, languages, and accents. These are commonplace examples of machine learning and artificial intelligence (AI). Many other impactful stories emerge when applied to medicine (e.g., imaging and diagnosis), business transactions, complex climate models, autonomous vehicles, and more. The rapid growth is fueled by low-cost, large-scale computing power and ubiquitous connectivity. Yet, beyond the benefits we receive as consumers, what additional factors should we pursue as environmental health professionals and data managers?

Artificial Intelligence

The term AI covers a long list of disciplines that make machines smarter (or make machines *seem* smarter). AI incorporates concepts such as machine learning, deep learning, natural language processing, image processing, and automated speech recognition. It takes a data scientist to understand it all but we can learn.

Many AI applications begin by training the system to observe previous experiences, either in real time or by mining historical data. Consider how a game of chess can be broken-down into elemental first, second, and third moves. Moves and countermoves alter the outcome likelihoods and by observing many outcomes (i.e., wins and losses), the system draws conclusions about how those elemental decisions impacted the outcomes. The accumulation of data can be packaged and presented as intelligence and helps inform what chess moves are likely to result in a win.

While Google, Amazon Web Services, and others provide AI platforms on a low-cost or pay-as-you-go basis in the cloud, for our purposes in this column, we will experiment with the Microsoft Azure AI Platform and follow a simple tutorial to understand how machine learning could apply to environmental health.

Considering Our Data

To look at hand washing and find patterns spanning many years of routine inspections, we are invited to select features surrounding each observation (Figure 1). So, for each IN and OUT occurrence, what might help predict future violations? Inspection/violation history, facility type, type of ownership, and food handler certifications are examples.

Create a Model

The model is trained by classifying our data in the hope that it can accurately predict

FIGURE **1**

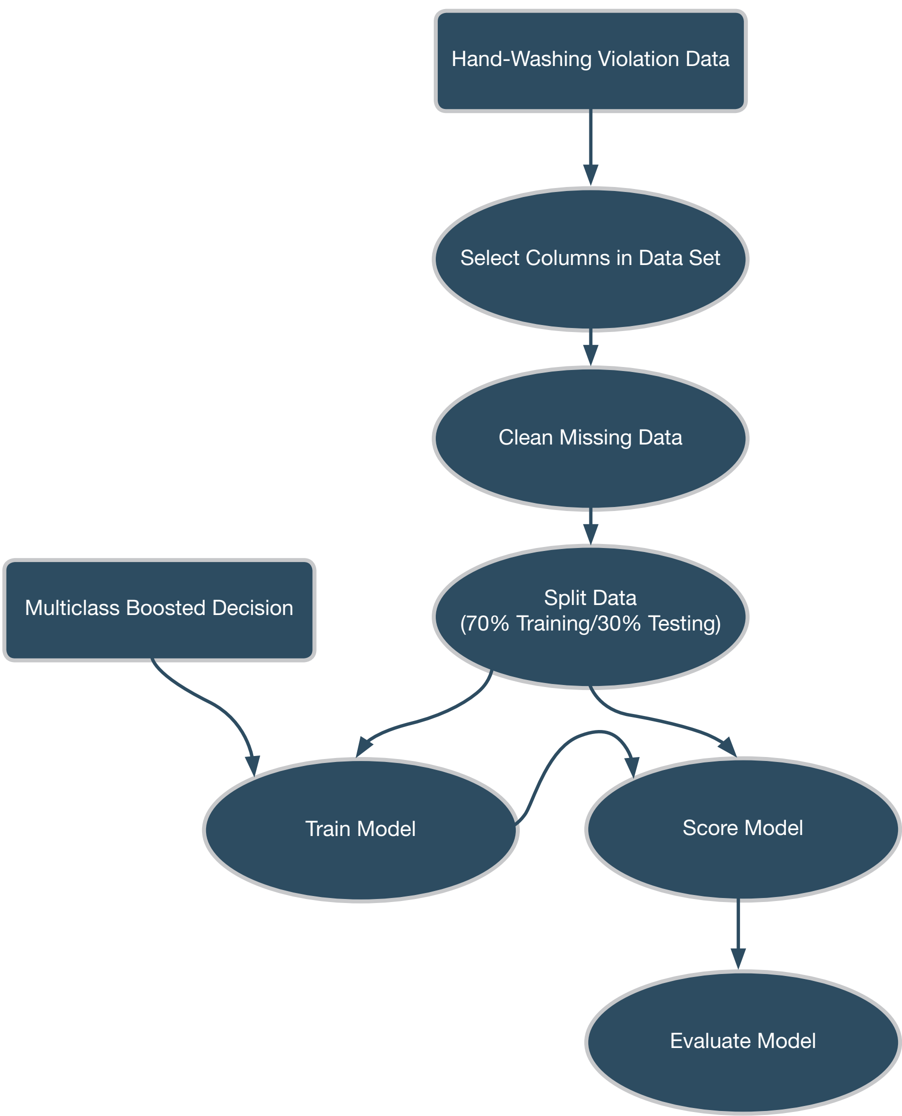
Example of a Hand-Washing Data Set

	A	B	C	D	E	F	G	H
1	Handwashing Observation	Assessed Risk	Facility Type	Ownership Type	Food Handler Cert	Prior Violation	Years in Related	Business Complaints
2	IN	Medium	Limited Prep	Corporate	Yes	No	12	No
3	IN	Medium	Full Service	Corporate	Yes	Yes	5	No
4	OUT	Medium	Full Service	Partnership	Yes	Yes	6	No
5	OUT	Medium	Full Service	Partnership	Yes	Yes	6	No

Note. Column A indicates an observed historical inspection result. Columns B–H are features intended to try to train a model.

FIGURE **2**

Artificial Intelligence Model



future outcomes. We've collected hundreds of thousands of observations related to hand washing. Can the model reliably predict future hand-washing issues using the features we selected?

Upon creating my free account and following a basic tutorial, I selected a feature that checks multiple machine learning methods. Since I don't know the method that fits our problem best, I let the system run them all—we'll see which is best.

Train and Test

I clicked "Run" to train the system. It took several minutes to run, train, and test the model (Figure 2). During this phase, the system consumed my data, split the streams, performed classification analyses, and tested the model.

I found the split data concept to be the most interesting. I learned that a common practice is to divide your data, retain a portion to train the model, and use the remaining to test the results. If the predictive model matches the test data, we gain confidence in our model. We used 70% of our data to train the model and 30% of our data to test the model.

When the process finished, I reviewed the results, which were a statistical analysis of our data and a score. The results and graphs indicated the degree to which our training data conclusions matched our sample data. With the portion of the data we fed into the system for training, could we predict outcomes in the sample of data we set aside for testing? If not, something went wrong and the model should be revised and rerun.

A primary learning point is that training models are an iterative process. A project might have many models and many experiments. After the model is selected, it should be refreshed and retrained over time.

Deploy and Utilize

Once satisfied that our model is valid, the model can be packaged and published in the cloud. A published model can be accessed in real time by other software systems such as websites or your own inspection system. So, in the same fashion an investor pulls up a stock price on the Internet or a family member pulls up weather forecasts, our simple model could be used to lookup (or predict) a routine inspection violation profile.

What would it mean to health department resourcing to predict likely violations as an understanding of facility risk? How would this information impact inspection frequencies, fees, and staffing?

Bias in the Model

It is easy to inject inappropriate biases into our models. Take a moment to be thoughtful about the features of the data set in the training exercise. To imply a gender, race, or economic standing into the model might be incorrectly emphasized or amplify related biases over time.

Conclusion and Next Steps

The model described above is simple and intended to be only thought-provoking. The computing resources and tutorials to repeat the basic example above are freely available to enthusiasts like me. That's a welcome democratization of cloud computing power.

A more meaningful and valid exploration of the capabilities should be executed by food safety experts and data scientists, pursing a consensus model that is sufficiently valid as to impact our day-to-day practices. A quick search of scholarly papers shows many promising projects and results published by highly qualified experts.

Resource

Microsoft Azure Artificial Intelligence Tutorial: https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-designer-automobile-price-train-score

Corresponding Author: Darryl Booth, General Manager, Environmental Health, Accela, 2633 Camino Ramon #500, San Ramon, CA 94583. E-mail: dbooth@accela.com.