



Web site address www.DAMAIndiana.org
LinkedIn Group DAMA Indiana

Facebook DAMA Indiana
Twitter @DAMAIndiana

Welcome to DAMA Indiana News!

Welcome to the autumn issue of the DAMA Indiana newsletter. In this issue, guest author David Schwab talks about Artificial Intelligence. Autumn is almost here and we are looking ahead to our next chapter meeting! The next meeting will be **Thursday, November 1, 2018**. Be sure to save the date!

Do you have a question for the board or fellow members? Are you wondering about future events? Have an idea for a newsletter article? The top of the newsletter contains the web site and social media contact information for the chapter. We would love to hear from you!

From the *President's* Editor's Pen...

By Christi Denney

Hello DAMA friends,

Surprise! Occasionally the Editor gets to take over this space and share the things that are on her mind. It is just one of the many perks of the job!

First, we would like to extend our thoughts and prayers to all of those who have been suffering through the hurricanes and devastating flooding in the US and Kerala. Our hearts go out to all of those affected. I am confident that we will again reach out to show compassion to our neighbors, both near and far!

The cool air has not yet arrived (why is it still 90 degrees in September???), but autumn is quickly approaching. Although I want to be thinking about apples, pumpkins, and hayrides, I imagine that many of you are thinking about the same thing that is on my mind - business planning for 2019. Is it just me or is anyone else still trying to plan out the remainder of 2018? How can we possibly determine the needs of 2019 when 2018 is still in flux? But somehow we do. Every year.

One item of interest for many will be data lakes. Several organizations have implemented data lakes, so talking about data lakes is hardly a "new" concept. But I have to wonder whether the existing data lakes have been built without the necessary architecture, governance, and metadata management strategies.

Continued on next page

INSIDE THIS ISSUE	
1	Welcome
1	President's Pen
2	Announcements
3	Elections
4	November Meeting
5	Guest Author
7	Professional Development
8	Contacts

President's Editor's Pen, cont.

More and more articles are being written on how not to create a data swamp. The data lake was even compared to a [junk drawer](#) in one instance. Neither sounds like a desirable future state for a data lake initiative.

One of the selling points of the data lake was that you could throw anything in there and you didn't have to force the content to conform to standards, such as those required by the data warehouse. Unfortunately, some of those standards provide a basic level of governance and oversight that is relevant to any type of content - even content in a data lake. The concept of governance may not be on the minds of those implementing the lake. Perhaps it is viewed as "overhead." Since anything can be thrown into the lake, without having to conform to a standard, governance may be viewed as unnecessary. But a lack of governance can result in a data swamp - an area filled with content that may not be relevant, understood, or easily found.

In an article for Dataversity (<http://www.dataversity.net/data-lake-vs-data-swamp-leveraging-enterprise-data/>), Michelle Knight highlights several articles that provide guidance when creating or maintaining data lakes. I encourage anyone who is thinking about a data lake to take a look at some of the resources on Dataversity.

We have a great chapter meeting planned for November, filled with a variety of great topics. I hope to see you there!

- Christi



Save the Date!

Our next DAMA meeting will be **Thursday, November 1, 2018** at CNO Financial in Carmel, Indiana.

Look for an agenda and location details in this issue.



DAMA Indiana Elections

At the October DAMA Indiana Meeting, we will be presenting the slate and voting on board member positions for 2019. Current members who are present during the October meeting are eligible to vote.

Please contact Diane Michael at vpadministration@damaindiana.org if you are interested in a board position or would like additional information.



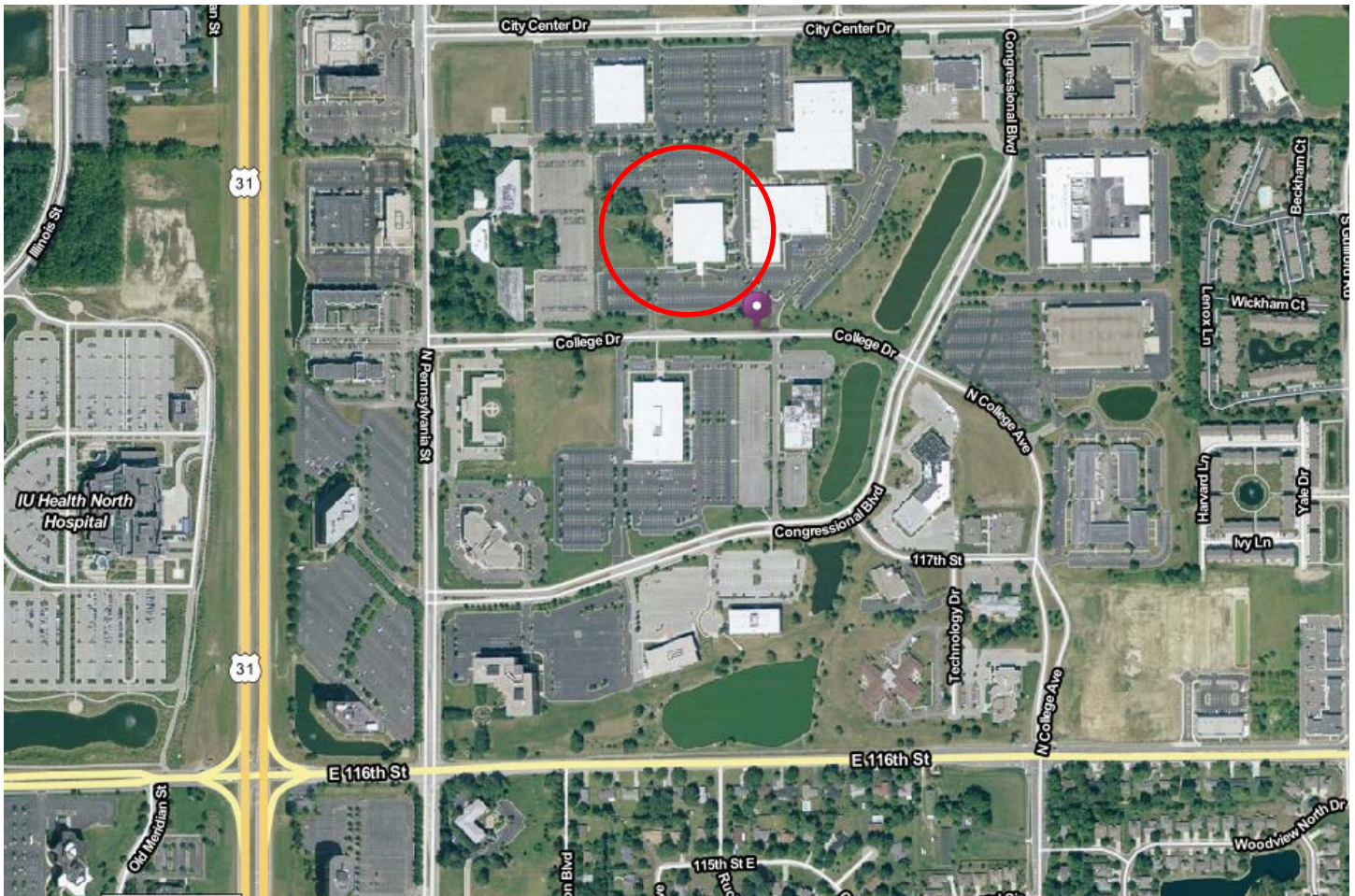
Save the Date - November Chapter Meeting

Join us on **Thursday, November 1st** for an informative day you won't want to miss! Our featured speaker for the day will be **Anthony Algin**, who will provide us with Data Leadership Lessons for Data Governance. Our own chapter member **David Schwab** will also be presenting an overview of Hadoop. **Moser Consulting** will be our sponsor for the event, and we will hear from **Warren Sifre** (Principal Consultant at Moser) on the topic of BI Architecture. We will also hold annual elections at this meeting. Continental Breakfast begins at 9:00am and we will wrap-up shortly after 3:00pm.

CNO Financial Group has graciously agreed to host this meeting. The address is **530 North College Drive, Building J, Carmel, IN 46032**. College Drive can be accessed from North Pennsylvania Street or East 116th Street. Park in any open, unreserved space in-front of the building. Below is a map to the CNO Financial Group's location.

Meeting information will be available on our website: www.damaindiana.org

Please register by sending an e-mail to info@damaindiana.org. We hope to see you there!



Artificial Intelligence Models

By David Schwab

Artificial intelligence models using machine learning have become a popular way to analyze large amounts of data. These models can find patterns that humans overlook, such as indicators of insurance fraud or similarities between customers. And with companies like Google and Amazon investing heavily in research, the potential benefits of machine learning models will only increase.

However, as the models have improved they have also become more difficult to understand and interpret. This isn't an issue for certain tasks, such as filtering e-mail. In other cases, however, the stakes are much higher: the majority of states use machine learning models to help sentence convicts, or to set bail (O'Neil, 30). Companies like Walmart, McDonalds, and Kroger use them to determine who has the appropriate "personality" for the job. We demand accountability when people make such decisions, and we must demand it when machines do as well.

This is easier for some types of models than for others. Consider the task of flagging health insurance claims for further review. We have a data set containing a large number of claims, and for each claim we know such things as the claim amount, the type of injury, and the number of claims the claimant has previously filed. We also know whether each claim was found to be fraudulent. Here are three different ways to approach this task.

First, we could use an error-based approach such as logistic regression. Error-based models minimize the total error between the model's predictions and the known outcomes. In this case, our model would tell us which features (e.g. claim amount, etc.) led to the greatest reduction in error; for example, we might find the probability of a claim being fraudulent increased by

1% for each dollar amount over \$5,000 (holding all else equal). This type of model can be highly accountable, because both the features used in the prediction and their relative contribution to the prediction are available.

We could also use a similarity-based approach such as k-nearest neighbor. These models classify an input based on how similar it is to known instances in the data set. The model can be designed to list the data points that it used for classification; therefore, we can say that a claim deserves investigation because it is similar to these 15 other claims, 9 of which are fraudulent. This type of model is not as accountable as the first one: we know the data points used to make the decision, but not why they were chosen instead of other data points, other than that the model said they were the most similar.¹

Both of these approaches assume the relationship between the features and the outcome is relatively simple. If that's not the case, we might need to use a more sophisticated model, such as a neural network. A neural network is a collection of interconnected nodes. Each node has a value assigned to it during the training process. When the network analyzes a claim, some of the nodes will be activated and add their contribution to the output. Of the three models presented here, this is the least accountable: we can't explain why certain nodes were activated, and the nodes themselves don't map to the features of the data set.

Part of holding machine learning models accountable involves selecting algorithms such as logistic regression when the data allows for it. When that's not the case, accountability must shift outside the model to the wider community using the model and affected by the outcomes it generates. Ideally, there should be an

¹ The details depend on the algorithm used: one popular approach for continuous features is to minimize the distance between the feature values of the test case and the values in the data set.

appeal process external to the model. This is not as far-fetched as it may sound: in Indiana, property tax assessments are initially calculated from an intricate set of cost schedules, which most taxpayers will not understand. However, a taxpayer can appeal his assessment by offering evidence of his property's market value, such as a recent sale of similar property. Likewise, a defendant who has been denied bail based on a machine learning model could show her ties to the local community, or other pertinent factors.

This appeals process is particularly vital when the model itself doesn't (or can't) consider these other factors: for example, a parole model might consider that the convict has been in four serious fights while incarcerated, but it probably won't know why the fights

occurred, or who the aggressor was. This type of "soft" data is not only difficult to include in a model, it is generally better evaluated by people than by machines.

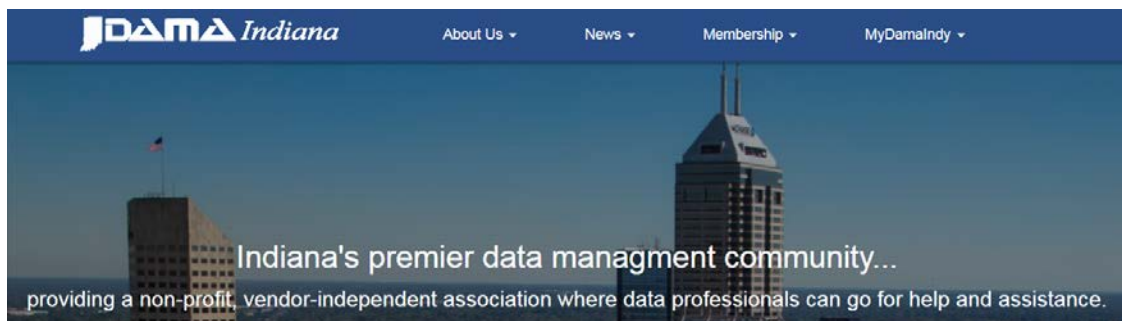
As always, the time to address model accountability is in the design stage. Just as reliability and security are built in from the beginning, every model should have an accountability plan that specifies both those affected by the model and their appeal options. Affording model accountability the respect it deserves will be key to ensuring that these powerful tools are kept in check.

References

O'Neil, Cathy. *Weapons of Math Destruction*. Broadway Books: New York, 2017.

Editor's Note: Thank you, David, for sharing your insights on Artificial Intelligence!

COMING SOON: a new and improved DAMA Indiana website!



Development work continues on our new DAMA Indiana web site.

Thanks to Adam, our awesome webmaster!

The initial release of the site is targeted for completion by the end of 2018.

Over time, we will incorporate new functionality, such as online event registration, online payments, and members-only content.

Professional Development Opportunities

THE TRANSFORMATION TO DATA-DRIVEN BUSINESS STARTS HERE

ENTERPRISE DATA WORLD



The next conference will be:

March 17-22, 2019

<http://www.dataversity.net/enterprise-data-world-2019-boston-ma/>

<http://dama.org>



Check out <http://tdwi.org> for conferences

DAMA Chicago

DAMA Chicago holds six bi-monthly meetings and other events. Their web site is: <http://www.damachicago.org/>

Southwest Ohio Chapter of DAMA

Check out their web page at:
<https://swoc-dama.wildapricot.org/>

Central Ohio DAMA

The central Ohio chapter page:
<https://buckeyedama.org/>

Here are other area events that may be of interest to data professionals:

IndyPASS (Professional Association for SqlServer)

Events are updated on their web site.
Their main web site: <http://indypass.org/>

INOUG (Indiana Oracle Users Group)
Check inoug.org for more information.

Reminder: Attending conferences and professional meeting counts towards CBIP and CDMP recertification credits. Be sure to record these in your log!

Contacts

Have a question or need to contact a board member? Below are the names and e-mail addresses of board members

DAMA Indiana Board

President: Gene Boomer
President@damaindiana.org

VP Administration: Diane Michael
VPAdministration@damaindiana.org

VP Communications: Christi Denney
newsletter@damaindiana.org

VP Finance: Open
VPFinance@damaindiana.org

VP Online Content: Open
VPOnlineContent@damaindiana.org

VP Programs: Michael Irick
VPPrograms@damaindiana.org



Found a great educational site?

Share it with us!

E-mail newsletter@damaindiana.org