

# A Short Introduction to Big Data [Analytics]

## Harnessing Insights for Strategic Decision Making

Fadel M. Megahed, PhD

Endres Associate Professor  
Farmer School of Business  
Miami University

 @FadelMegahed

 fmegahed

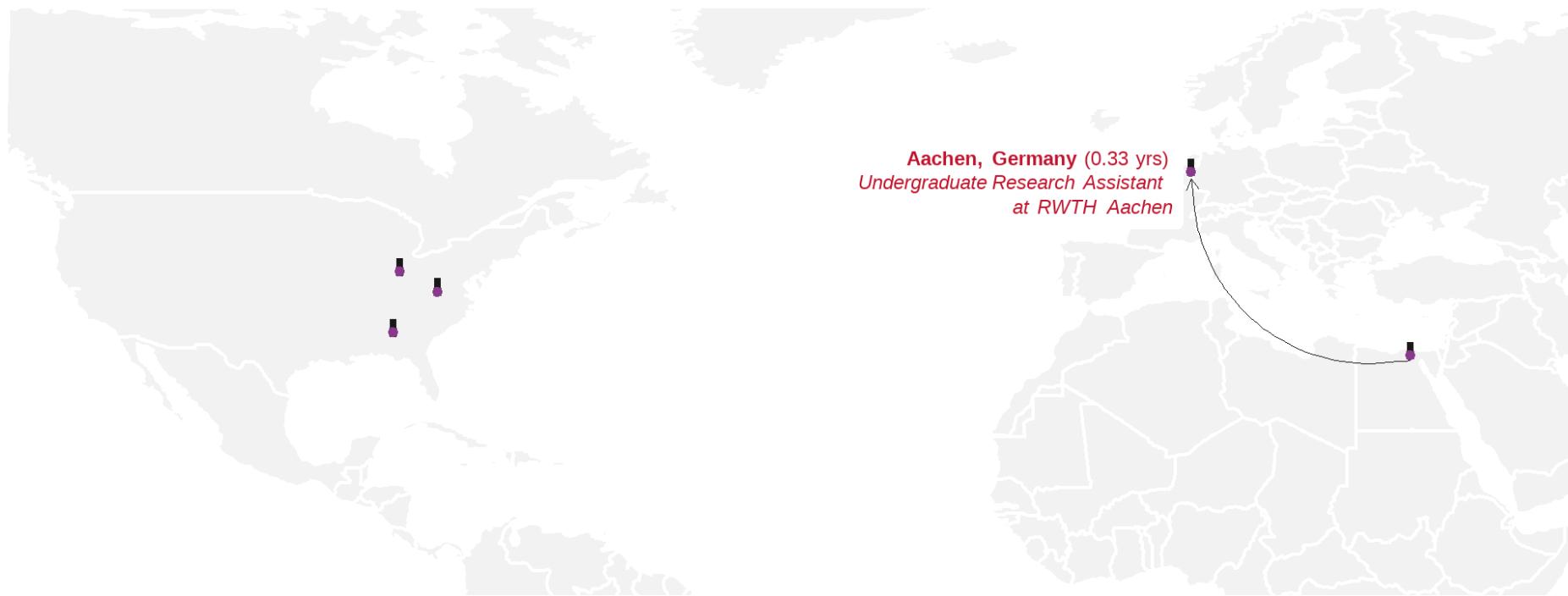
 fmegahed@miamioh.edu

Spring 2023

# About Me – My route to Miami University

- Application of data-driven decisions (D3) in 3 continents.
- **Interests:** Applications in logistics, manufacturing, occupational safety & portfolios.
- **Collaborations with:** Aflac, GE Research, Gore, Hilton, IBM Research, & Tennibot

Home 2 of 6



# About Miami University



# Learning Objectives for Today's Class

- Define **big data**, describe its **evolution**, and explain its importance.
- Provide examples of **big data sources**.
- Describe the **main steps and technologies/tools used** in **big data analytics**.
- Discuss the various **applications of big data analytics in different business sectors**.
- Discuss **future trends in big data**.

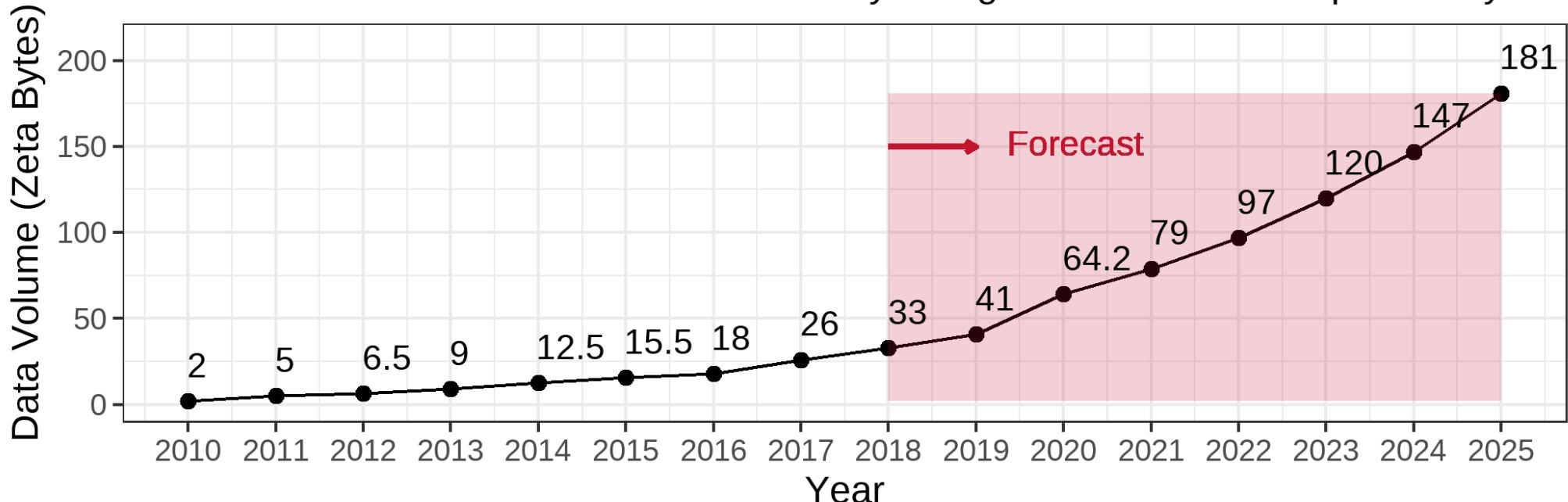
# What is Big Data?

# Big Data: A Working Definition

Big data is used to describe the **large, diverse, complex** and/or **longitudinal datasets** generated from **a variety of instruments, sensors and/or computer-based transactions**.

## Exponential Growth in Data Generated

Most of the data that exists in the world today was generated within the past five years.

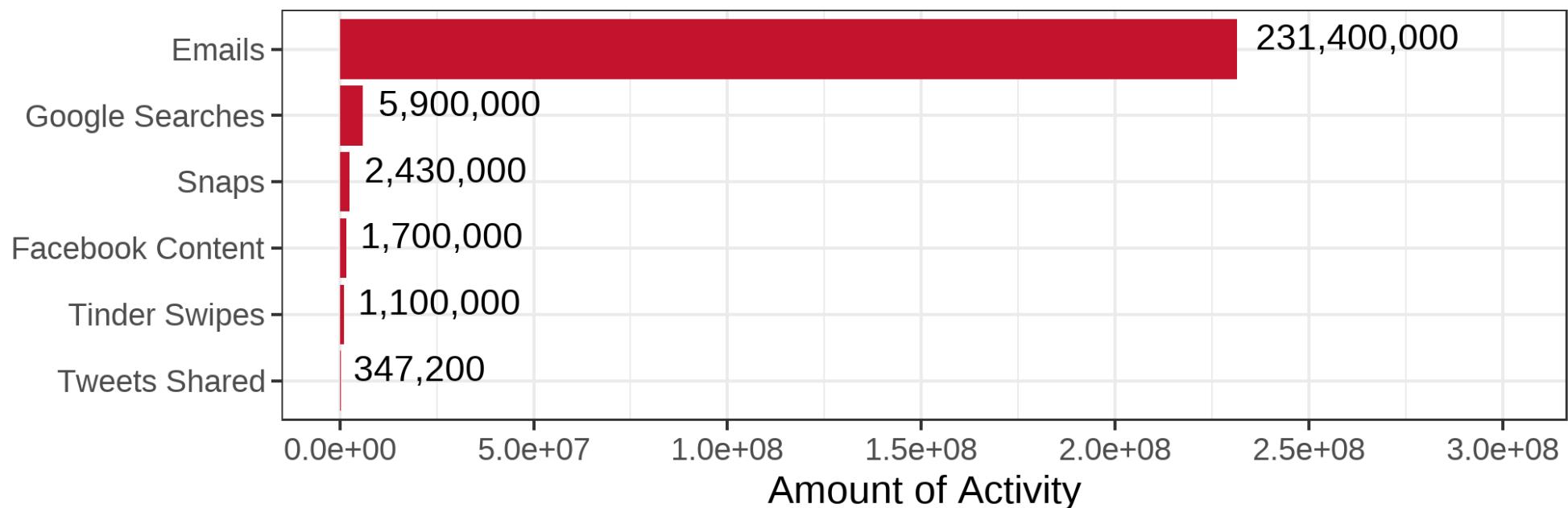


Credits: Created by Fadel M. Megahed | Data Source: Statistica

# Big Data: The Three Vs

The term big data refers not only to the size or **volume** of data, but also to the **variety** of data and the **velocity** or speed of data accrual.

Media usage in an internet minute as of April 2022  
The volume and velocity of data generated are very large!!!



Credits: Created by Fadel M. Megahed | Data Source: Statistica

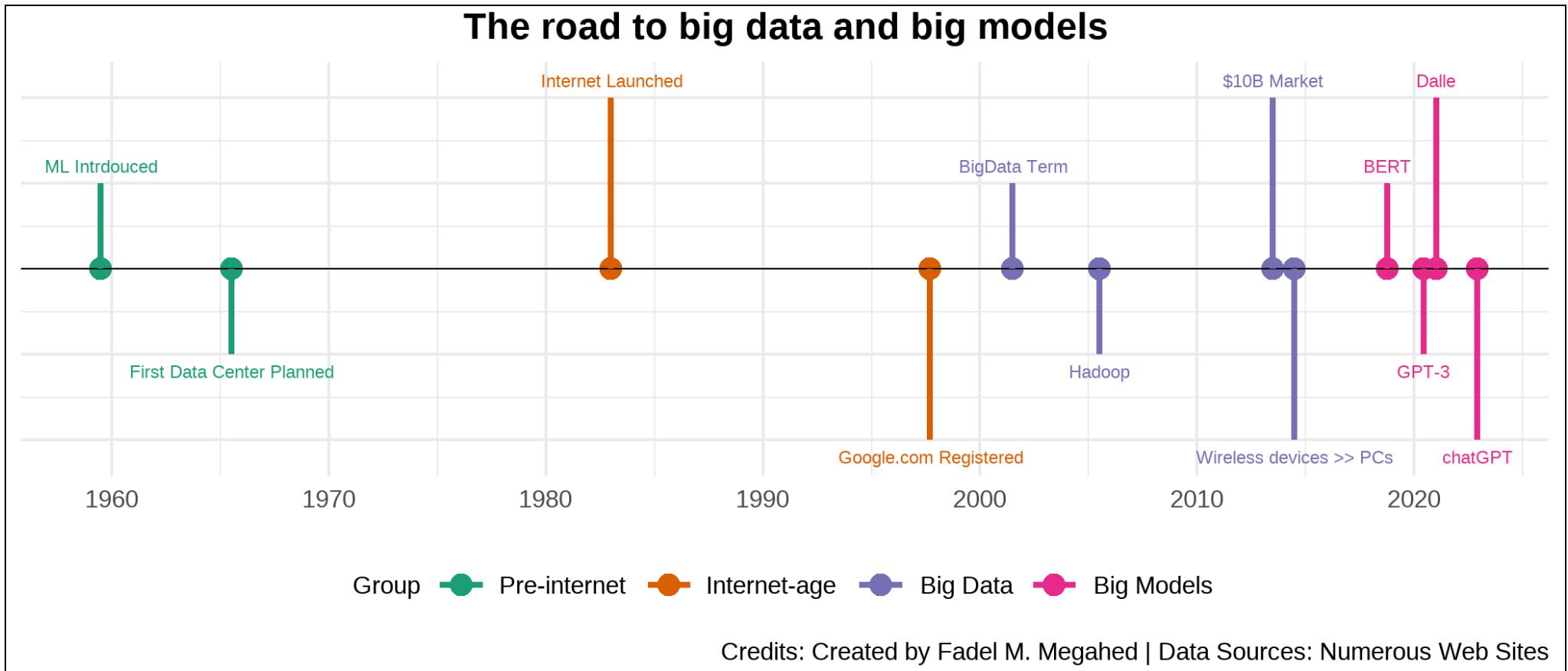
# Big Data: Beyond the 3 Vs

- The term big data refers not only to the size or **volume** of data, but also to the **variety** of data and the **velocity** or speed of data accrual.
- Other **V's** important to big data problems are:
  - **Veracity:** Trustworthiness of the data
  - **Value:** Added value to creating knowledge in a topic
- Both veracity and value relate to *data quality*



- The **infamous** hacking of the AP account and the associated fake tweet.
- The tweet caused the U.S. Dow Jones Index to drop momentarily by 140 points.

# The Evolution of Big Data



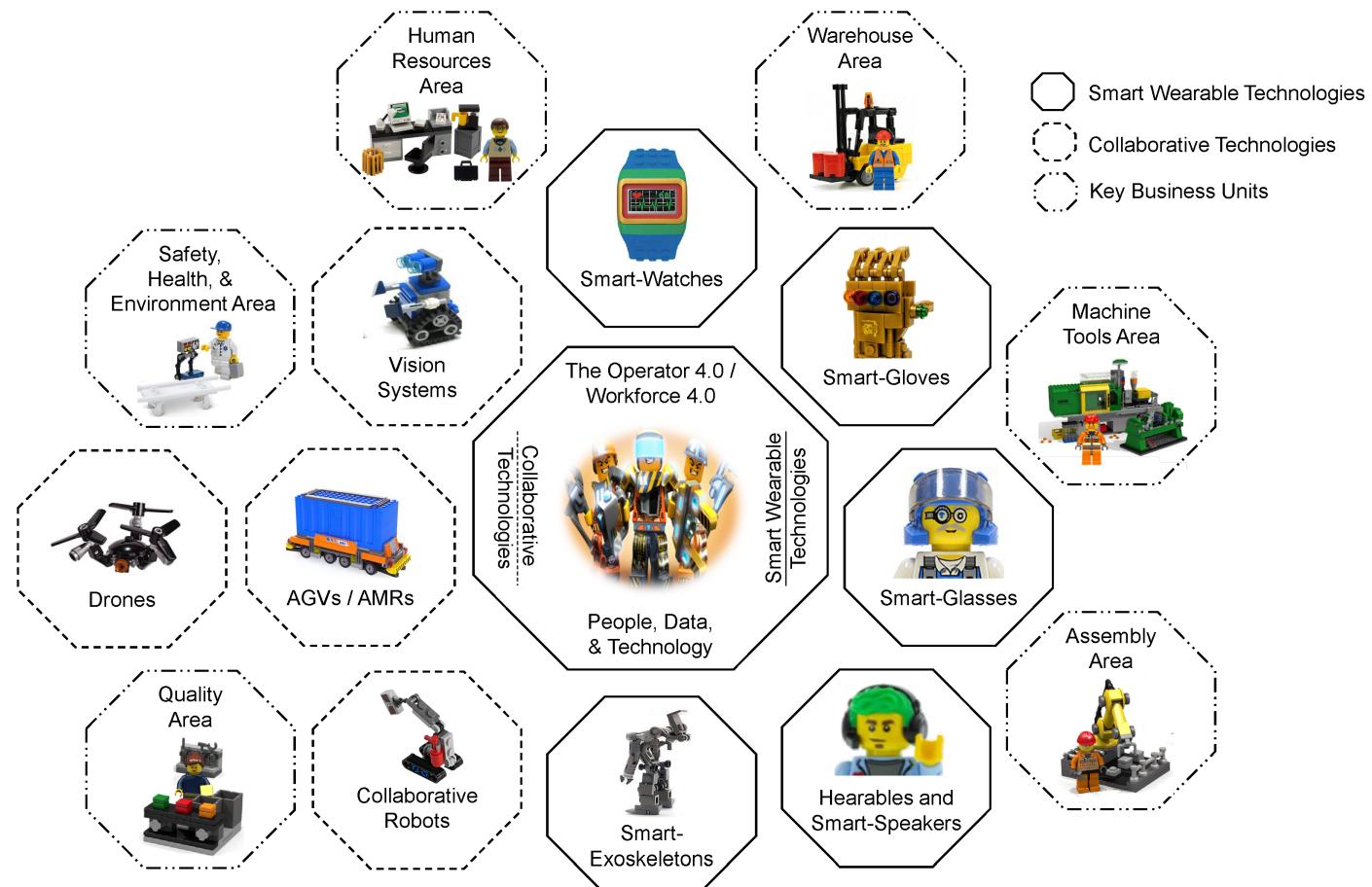
# A Tour of Big Data Sources

Examples from the Manufacturing and Transportation Industries

and

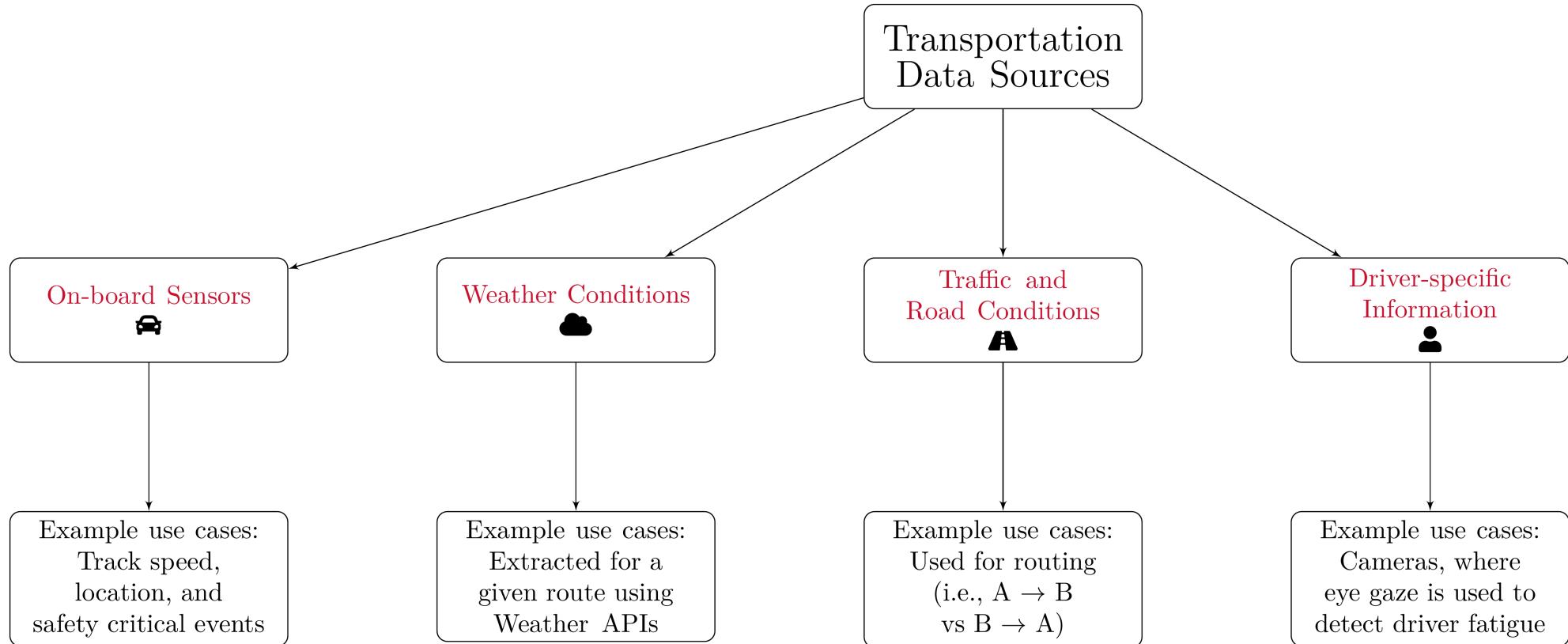
a discussion of the Healthcare/Medical Industry

# An Operator-Focused Manufacturing Example



**Source:** Figure is from D. Romero, T. Wuest, M. Keepers, L. A. Cavuoto, and **F. M. Megahed**, "Smart Wearable and Collaborative Technologies for the Operator 4.0 in the Present and Post-COVID Digital Manufacturing Worlds," *Smart and Sustainable Manufacturing Systems* 5, no. 1 (2021): 148–166. <https://doi.org/10.1520/SSMS20200084>.

# A Transportation Industry Example



Go to [www.menti.com/alpb24pop6](http://www.menti.com/alpb24pop6)

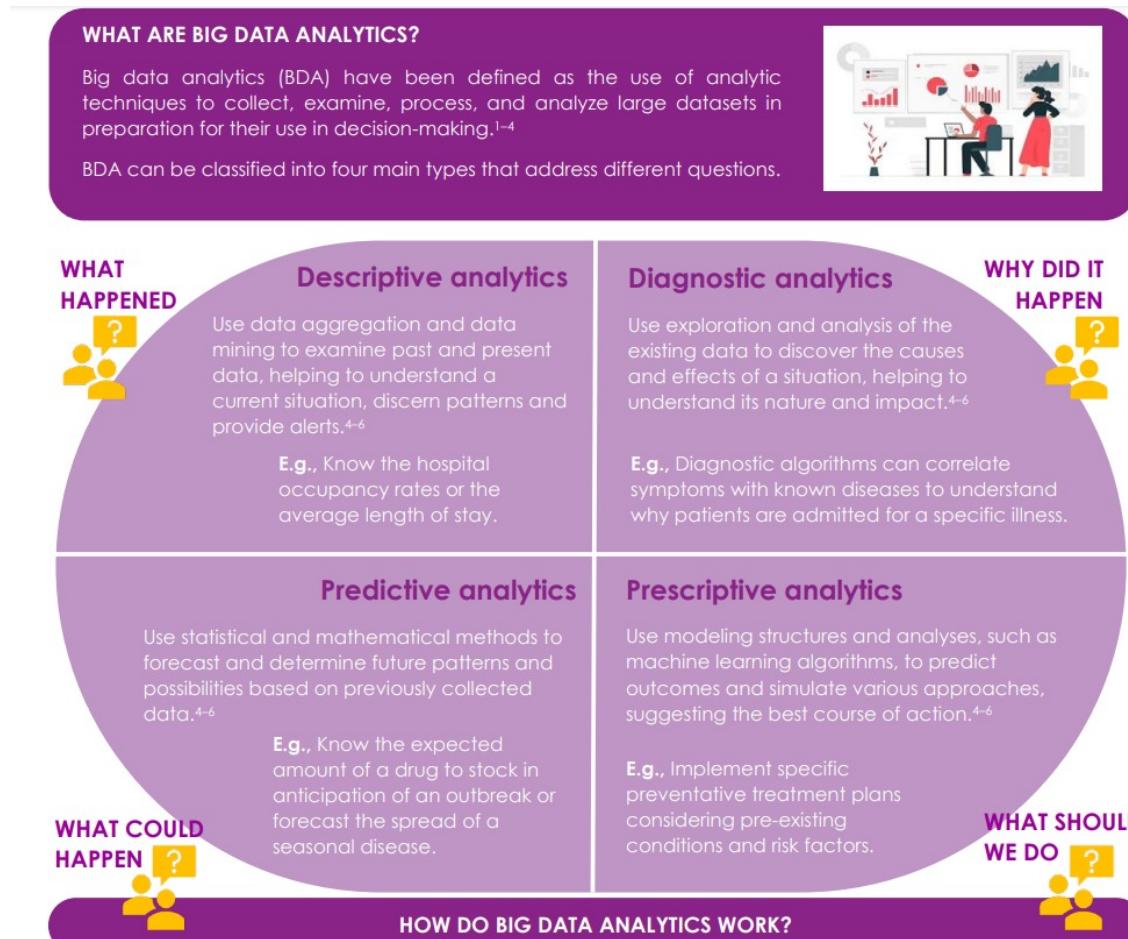
02 : 00

What are possible big data sources for the healthcare/medical industry?



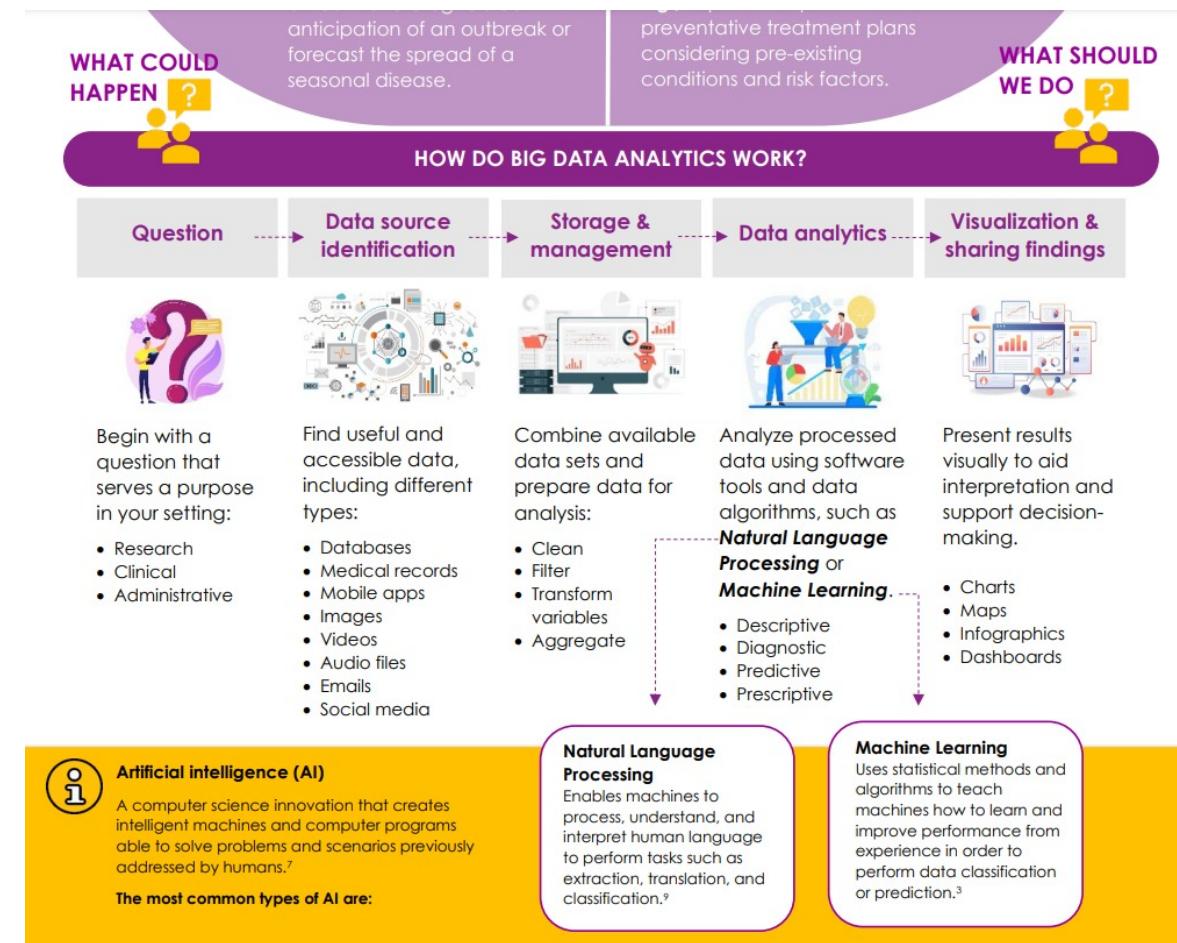
# What is Big Data Analytics About?

# Steps in Big (and Small) Data Analytics



**Image Source:** The image is part of an infographic created by Canada's National Collaborating Centre for Infectious Diseases, see [https://nccid.ca/wp-content/uploads/sites/2/2023/04/Big-data\\_Analytics\\_Apr2023.pdf](https://nccid.ca/wp-content/uploads/sites/2/2023/04/Big-data_Analytics_Apr2023.pdf).

# Steps in Big (and Small) Data Analytics

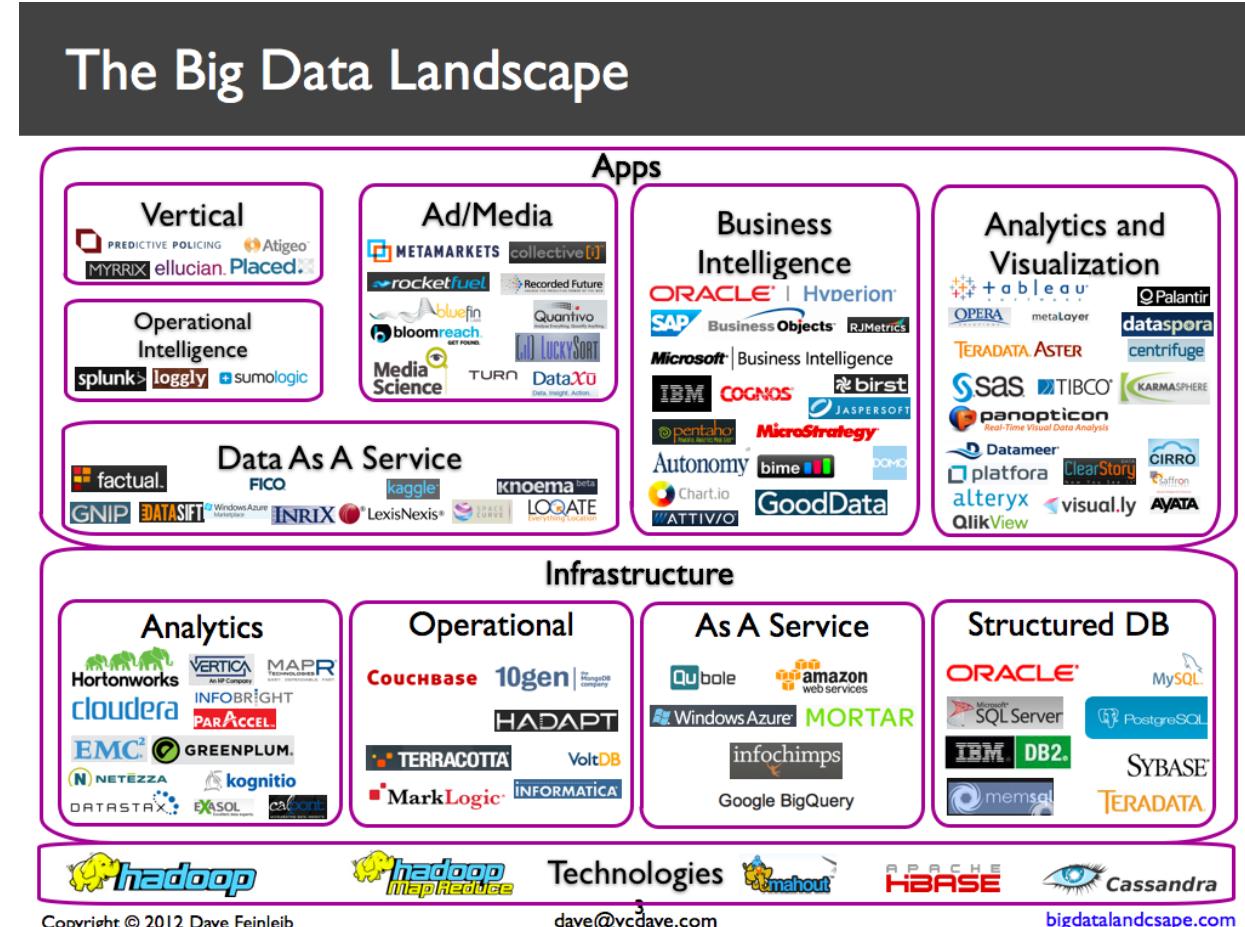


**Image Source:** The image is part of an infographic created by Canada's National Collaborating Centre for Infectious Diseases, see [https://nccid.ca/wp-content/uploads/sites/2/2023/04/Big-data\\_Analytics\\_Apr2023.pdf](https://nccid.ca/wp-content/uploads/sites/2/2023/04/Big-data_Analytics_Apr2023.pdf).

# Some Practical Advice for Big Data Analytics

- **Question the ?** Are you answering a question worth answering? **Before you start:** Attempt to quantify the return on investment (ROI) on the best possible case (perfect model).
- **Correct ✓** data, i.e.,
  - *Do you actually capture the important predictors?*
  - *Is your data aggregated to the right level?*
- **Cleaned 🍴** data, i.e.,
  - *Is your data worth analyzing?*
- **Iterate** ""Fail early, fail often, but always fail forward", **prior to deployment**. Also have time-points, where you can 🚧 the project if need to.

# Big Data Technologies (2000-2013)



# Big Data Technologies (What to Focus On?)

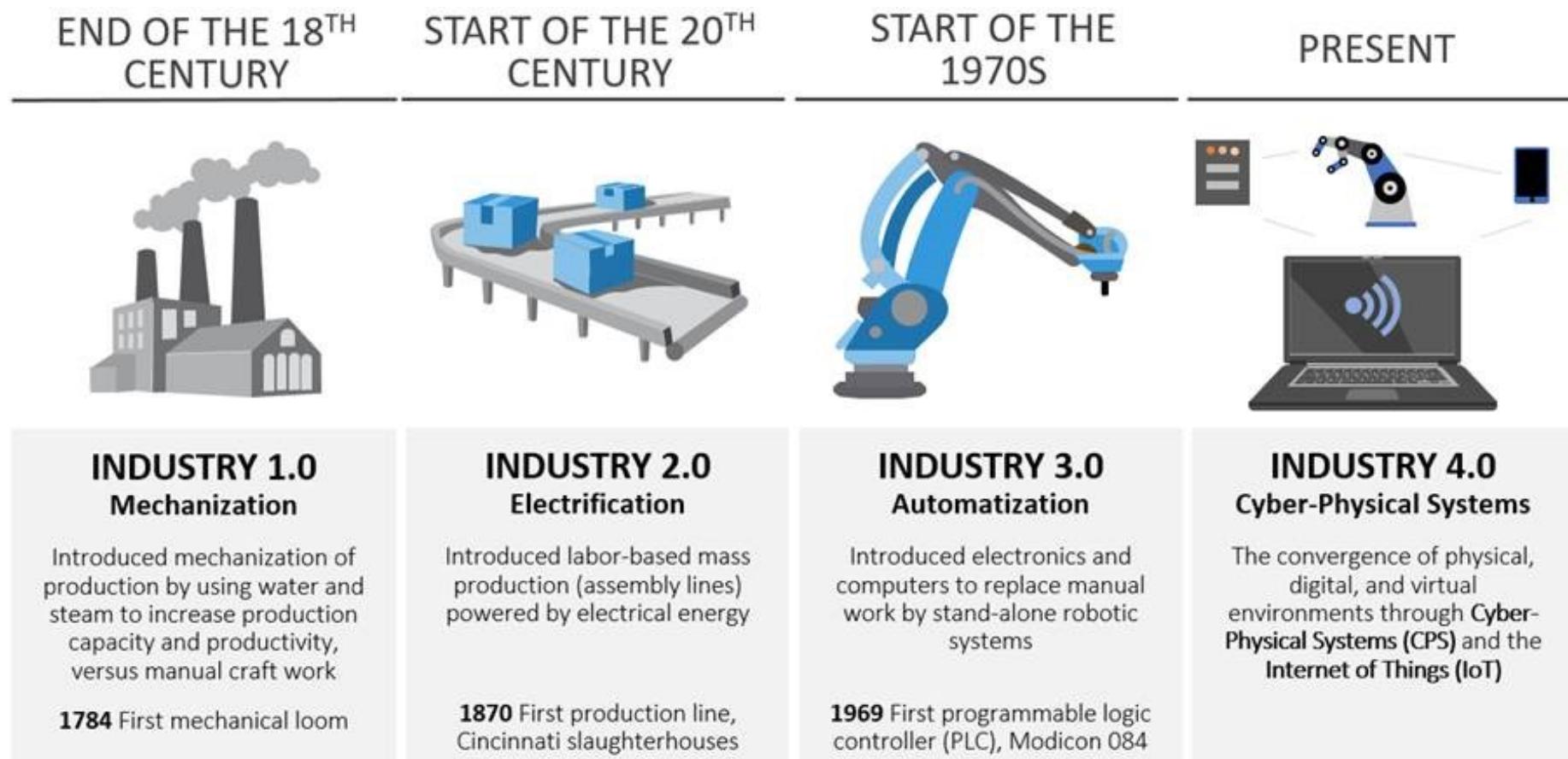
I do not argue against learning the basics behind specific technologies (e.g., using the excellent [MMDS book](#) as a starting reference), but as a business student, I would argue you should first focus on how to **ask important questions, assess data and solution needs, etc.**

Some **Disclaimers:**

- Technology is constantly changing and you will be expected to continue to learn and grow in your career.
- Hype and trendy topics change (e.g., [Gartner's new outlook on big data](#) is quite different from their 2010-2015 messaging), but the core problem remains the same, how to leverage internal and external data to drive value  for your organization.

# Example Big Data Applications

# Manufacturing 4.0



# Manufacturing 4.0



# Manufacturing 4.0: A Big Data Example

- **Goal:** Detect manufacturing defects in 3D printing (additive manufacturing) in near real-time.
- **Sensor Choice:** A thermal camera to detect anomalies in the cooling patterns for each-built layer.
- **Approach:** An image monitoring of the resulting thermal camera images.
- **Payoff:** Reduction of scrapped and wasted materials, improved **real** utilization of the 3d printers.

# Manufacturing 4.0: A Big Data Example

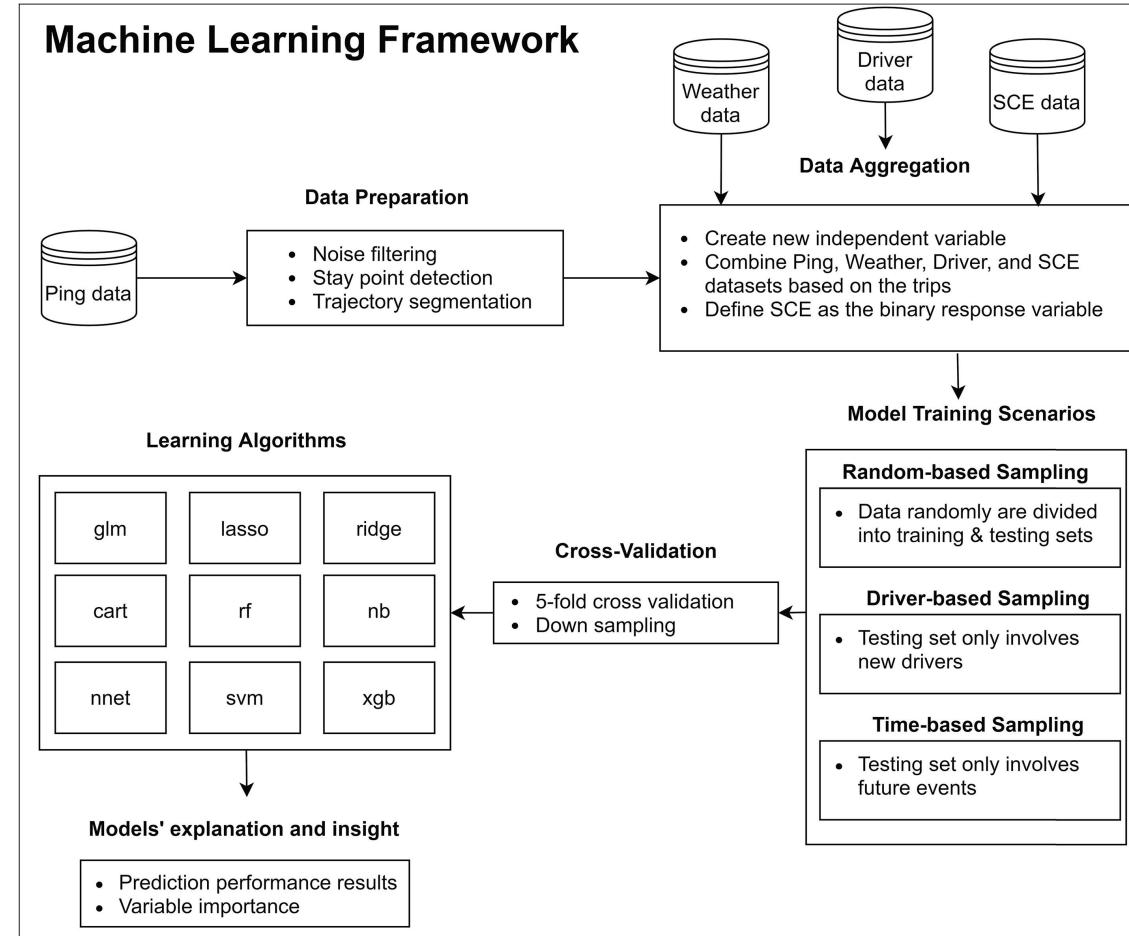


**Video Source:** How thermal cameras can be used to detect layer-by-layer additive manufacturing defects. Video combines work done at Auburn, with online thermal camera illustrations. Note that we use this illustration to highlight the use case of image monitoring for additive manufacturing, without exposing any real data from industrial partners.

# Trucking 4.0

- **Goal:** Examine whether driver-recorded safety critical events can be predicted (with a future 30-minute window used to allow for rest-break scheduling).
- **Sensor Choice:** A smart breaking system.
- **Approach:** A supervised machine learning approach was utilized to predict the occurrence (yes/no) of a safety critical event in the next 30-minute window, with features capturing *weather, mapping APIs* and *driver's trip history*.
- **Payoff:** Minimizing crash risk (which is beneficial from a human safety perspective as well as a monetary perspective to the trucking company).

# Trucking 4.0: A Big Data Example



- The occurrence of **safety critical events can be predicted in 30-minute intervals.**
- Developed machine learning models are **generalizable to new drivers** and do **not require frequent training.**
- **Improved safety outcomes can be achieved** by deploying our models as a part of existing routing and rest-break scheduling models.

# Recap

# Summary of Main Points

By now, you should be able to do the following:

- Define **big data**, describe its **evolution**, and explain its importance.
- Provide examples of **big data sources**.
- Describe the **main steps and technologies/tools used in big data analytics**.
- Discuss the various **applications of big data analytics in different business sectors**.
- Discuss **future trends in big data**.

# Questions

