

# "Modern Machine Learning Methods for Educational Research"

**Dr.Keith Zvoch**

Professor in the Department of Education Studies,  
College of Education, The University of Oregon (UO)

**Christopher M. Loan**

The Department of Education Studies,  
College of Education, The University of Oregon (UO)

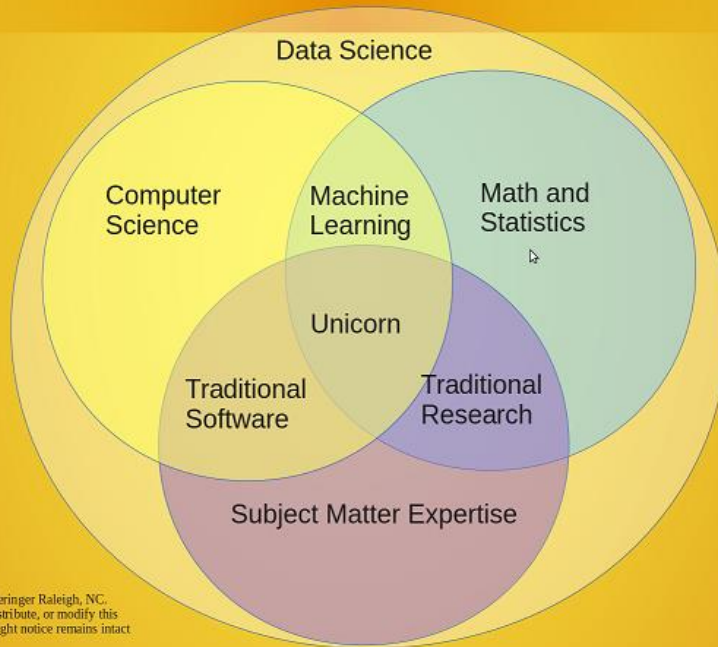


# Overview

- Introduction
- Decision tree & Random forest
- DS / ML ethics
- Data science research process
- Material / learning sources

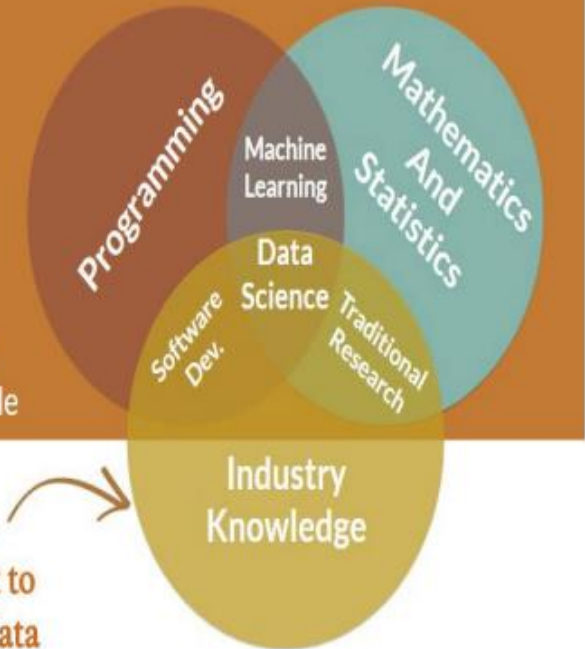
# Introduction

## Data Science Venn Diagram v2.0



Copyright © 2014 by Steven Geringer Raleigh, NC.  
Permission is granted to use, distribute, or modify this  
image, provided that this copyright notice remains intact.

- Data science applies the scientific method to analyzing data
- It lies at the intersection of several disciplines
- It draws on industry knowledge that makes the analysis of Big Data possible



Industry knowledge is  
essential to knowing what to  
look for when exploring data

DATA SOCIETY © 2021

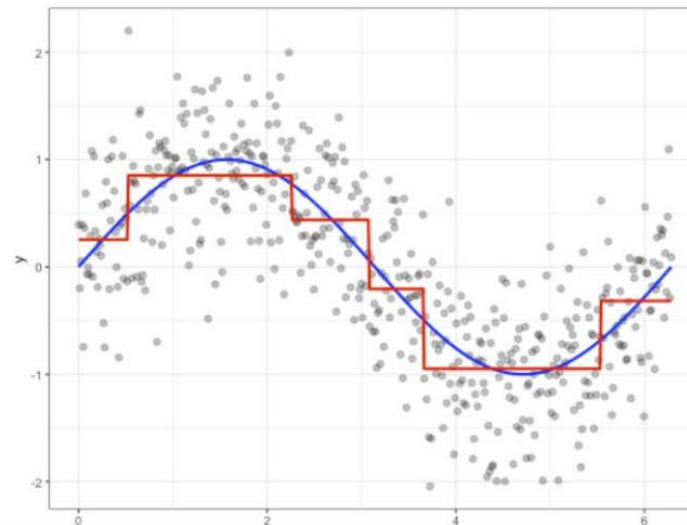
# Introduction

	Supervised Learning	Unsupervised Learning
Discrete	Classification: decision tree, random forest	Clustering
Continuous	Regression	Dimensionality reduction

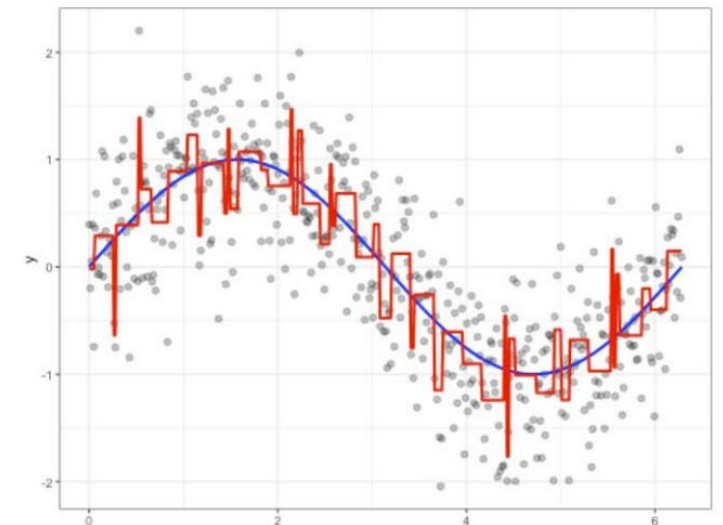
# Decision tree

- การแปลผลการวิเคราะห์และพัฒนาโมเดลได้ง่าย
  - ใช้ได้กับข้อมูลที่เป็น Cat. และ Cont.
  - ประยุกต์ใช้กับชุดข้อมูลที่มี missing data
  - ประยุกต์ใช้กับข้อมูลที่เป็น nonlinear relationships
- ❶ **Overfitting**
  - ❷ ประสิทธิภาพของการทำนายที่ลดลง เนื่องจากความแปรปรวนของข้อมูลที่มีน้อย หรือมีบางตัวแปรที่ dominate

Single pruned trees are poor predictors



Single deep trees are noisy



# Decision tree VS Random forest

## Decision tree

- ❶ Overfitting
- ❷ ประสิทธิภาพของการทำนายที่ลดลง เนื่องจากความแปรปรวนของข้อมูลที่มีน้อย หรือมีบางตัวแปรที่ dominate

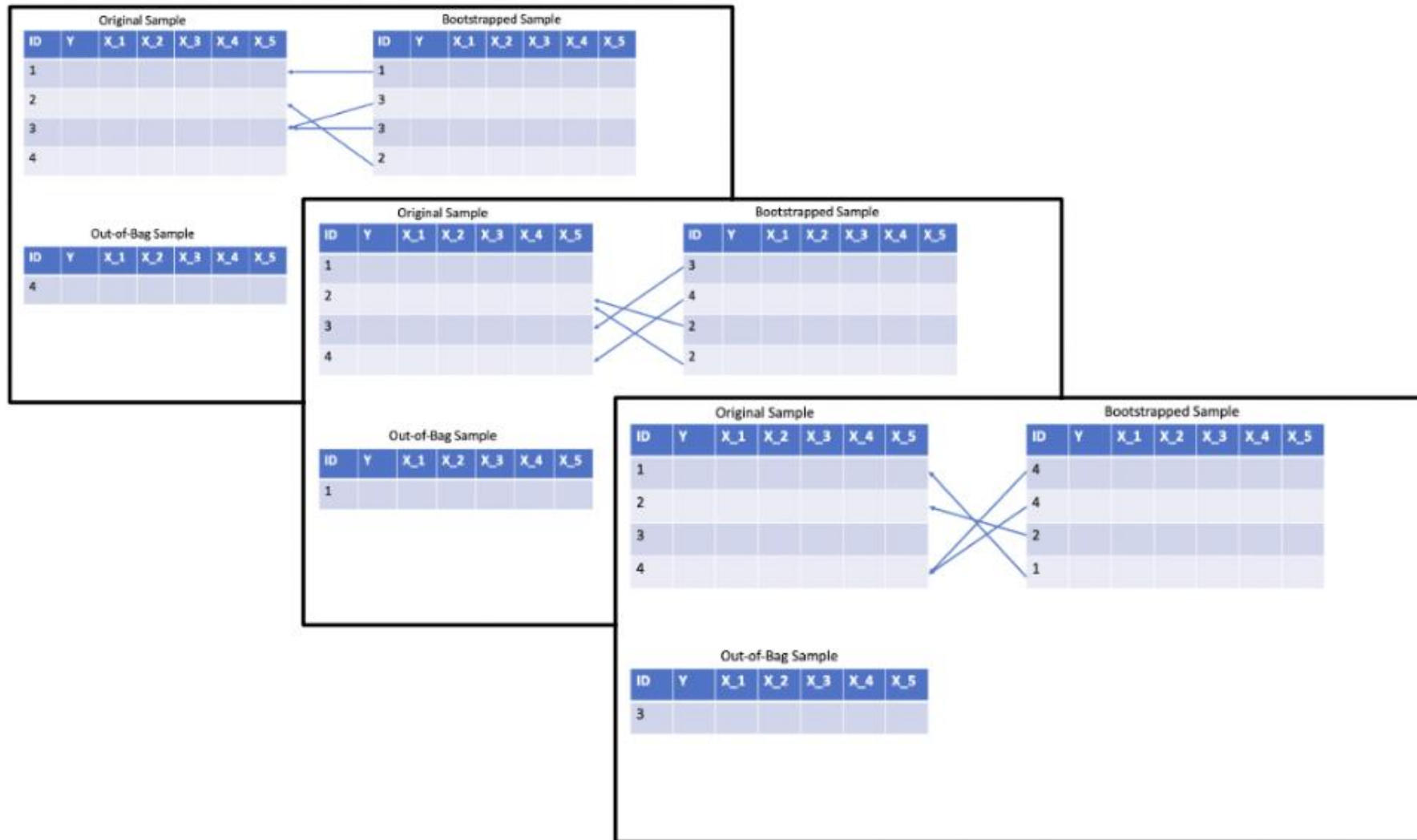


## Random forest

- การพัฒนาหลาย ๆ DT ด้วยข้อมูลจาก bootstrapping
- การเพิ่มประสิทธิภาพการทำนาย จากการทดสอบในทุก DT และได้ผลการทำนายด้วยวิธีการ voting

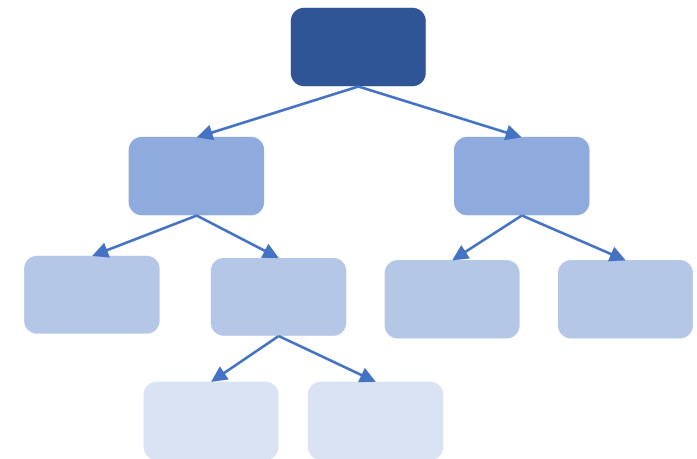
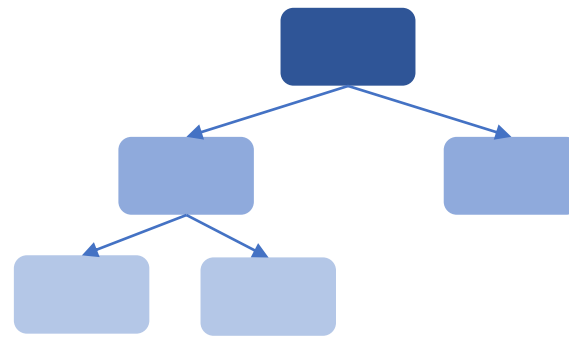
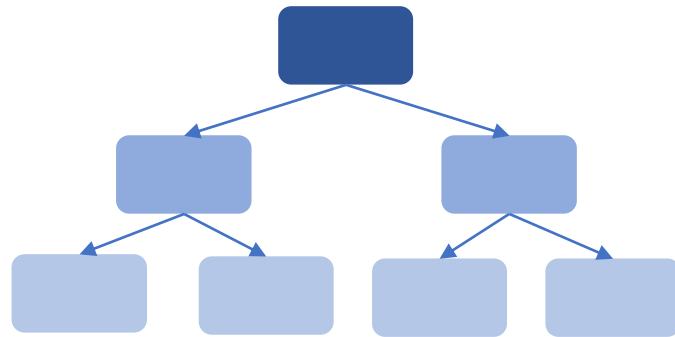


# Random forest



# Random forest

V1	V2	V3	outcome
xx	xx	xx	?



voting

Outcome	
Yes	No
2	1



เราใช้ DS/ML อย่างไรบ้าง



# Data Science Ethics

เราใช้ DS/ML อย่างไรบ้าง : การตัดสินใจ (decision making)

- การศึกษา (การเรียนรู้ต่อ, การทำนายผลการเรียน, ฯลฯ)
- ความบันเทิง (recommendation system ของ Netflix, YouTube, ...)
- การโฆษณาและประชาสัมพันธ์
- การตัดสินใจโทษทางกฎหมาย
- การจ้างงาน
- การแพทย์
- การเงิน

# Data Science Ethics

Bernard Parker

Prior offense:  
resisting arrest  
without violence

Subsequent arrests:  
None



High Risk: 10



Low Risk: 3

Dylan Fugett

Prior offense:  
attempted burglary  
with drug possession

Subsequent arrests:  
Three drug  
possessions

# Data Science Ethics

เราใช้ DS/ML อย่างไรบ้าง : การตัดสินใจ (decision making)

- โมเดล ML อาจให้ผลการวิเคราะห์ที่แสดงถึงความเอนเอียง (bias) ต่อกการตัดสินใจ
- การพัฒนาโมเดล ML ได้อย่างมีประสิทธิภาพ สามารถแสดงถึงความยุติธรรมในการช่วยตัดสินใจ (โดยเฉพาะในประเด็นอ่อนไหว เช่น ความเท่าเทียม ชนกลุ่มน้อย บุคคลที่มีความบกพร่องทางสุขภาพจิต)

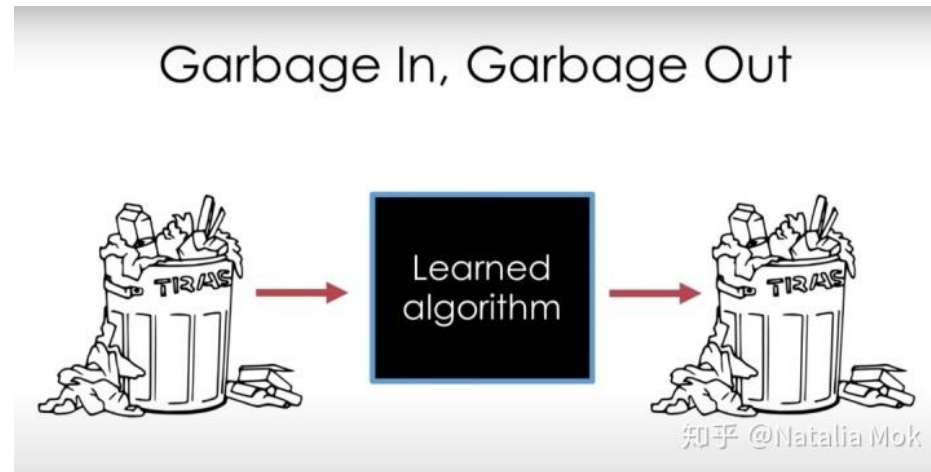
เราหลีกเลี่ยงได้อย่างไรบ้าง



# Data Science Ethics

เราหลีกเลี่ยงได้อย่างไรบ้าง :

- การให้ข้อมูลที่ปราศจากอคติ หรือการเลือกปฏิบัติ เนื่องจาก DS/ML ช่วยลดความเอนเอียงในการตัดสินใจของบุคคล โดยอิงตามข้อมูลที่มี
- การหลีกเลี่ยงการใช้ข้อมูลในอดีต หรือข้อมูลที่มีไม่สมบูรณ์



- การคำนึงถึงผลกระทบที่จะเกิดขึ้นจากการวิเคราะห์

# Data Science Ethics

## Principals of ethics :

1. Autonomy:

the right to control your own data

2. Informed consent:

the ability to explicitly approve use of your data based on understanding

3. Beneficence:

the use of your data should provide a benefit to you

4. Non-maleficence:

do not harm



# Data science research process

