



Are New Modeling Techniques Worth It?

Tom Zougas PhD PEng, Manager – Data Science, TransUnion

TORONTO SAS USER GROUP MAY 2, 2018



Are New Modeling Techniques Worth It?



Presenter

Tom Zougas PhD PEng, Manager – Data Science, TransUnion

As Senior Manager of Data Science at TransUnion Canada, Tom and his team are tasked with delivering advanced analytics projects.

He has 20 years of technical consulting experience in data analysis, system design, and software development. Tom has worked with clients in industries as diverse as financial services, insurance, software, utilities, telecom, pharmaceuticals and metals.

Prior to joining TransUnion, he was Director of Analytics at Angoss Software, where he managed a team of data scientists, worked as a senior advanced analytics consultant at IBM and at SAS, and managed the advanced analytics team at BlackBerry (RIM).

He has also authored and taught courses in data mining. Tom holds a PhD in Engineering from the University of Toronto.

Agenda



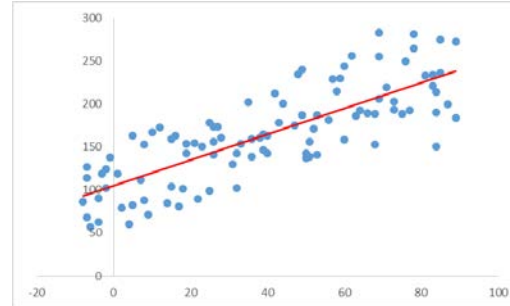
- Know The Core
 - Applying The Core Models
 - A Sampling of New Model Types
 - Q&A



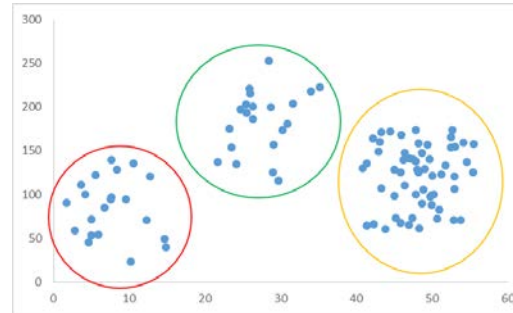
Know The Core

Top 3 Machine Learning Methods

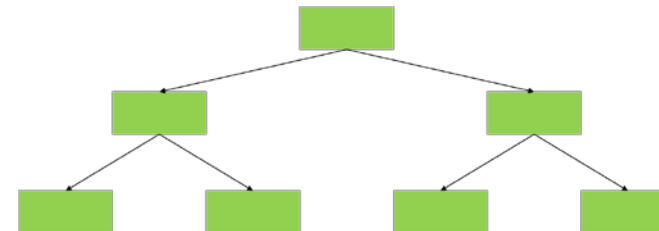
- Regression



- Clustering



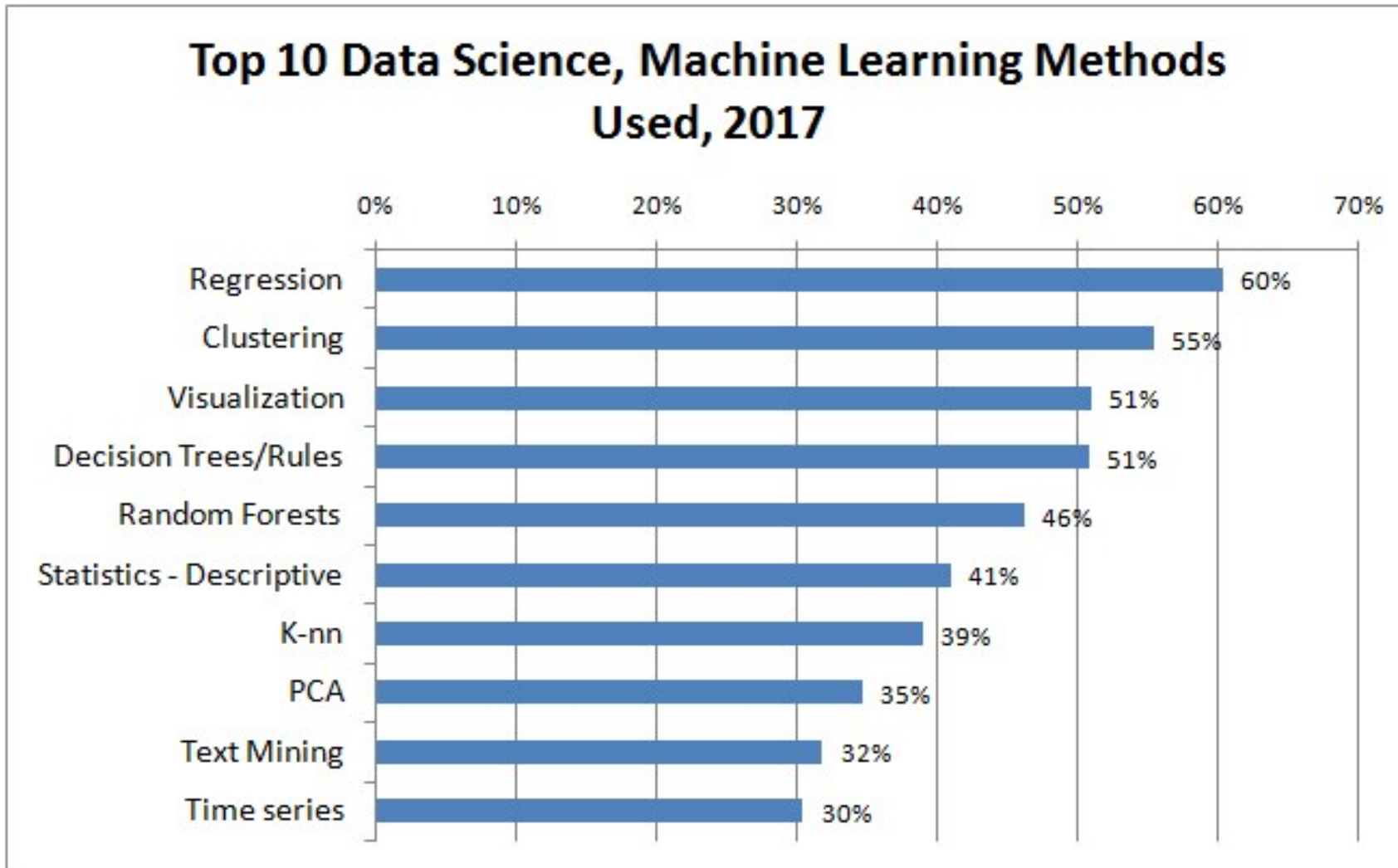
- Decision Trees



KDnuggets 2017 Survey



Source: <https://www.kdnuggets.com/2017/12/top-data-science-machine-learning-methods.html>



What Makes Them the Top 3

- Are they state of the art? ...No
- Are they leading edge? ...No
- Are they exotic? ...No
- What are they?
 - ✓ Simple
 - ✓ Interpretable
 - ✓ They Work

Categories of Machine Learning Algorithms

- Supervised Learning

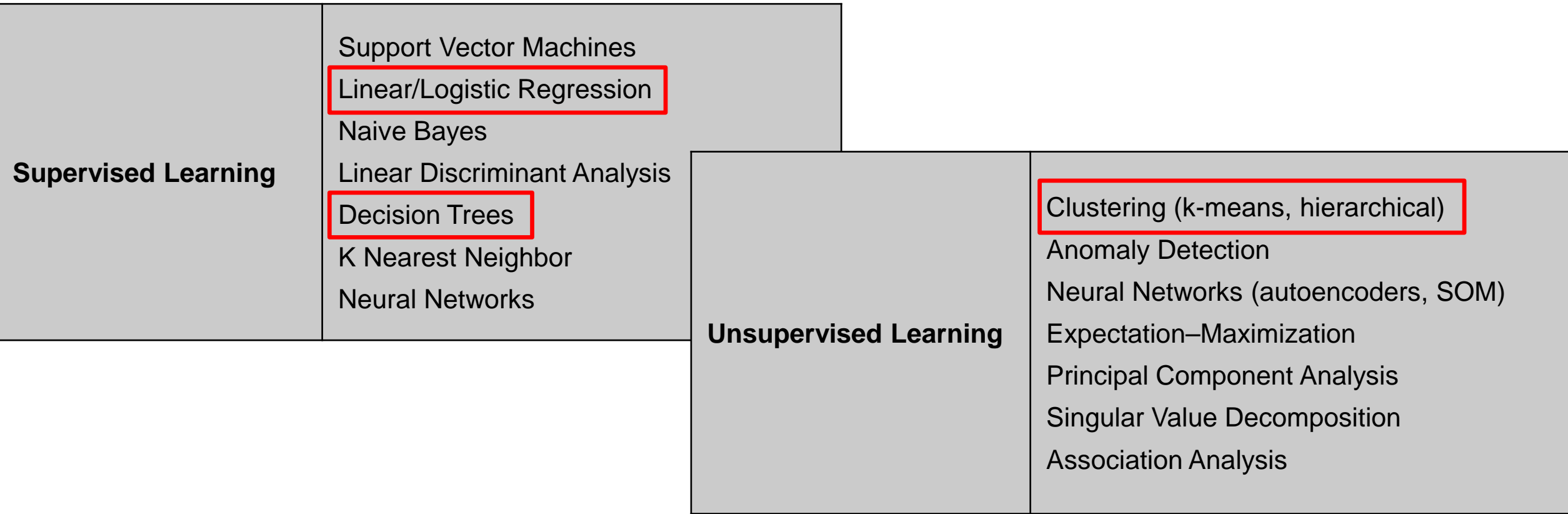


- Unsupervised Learning



- Others

Landscape of Machine Learning Algorithms



Often times, the baseline is sufficient to satisfy the business objective.



Applying The Core Models

What Analytics Methodology Do You Use



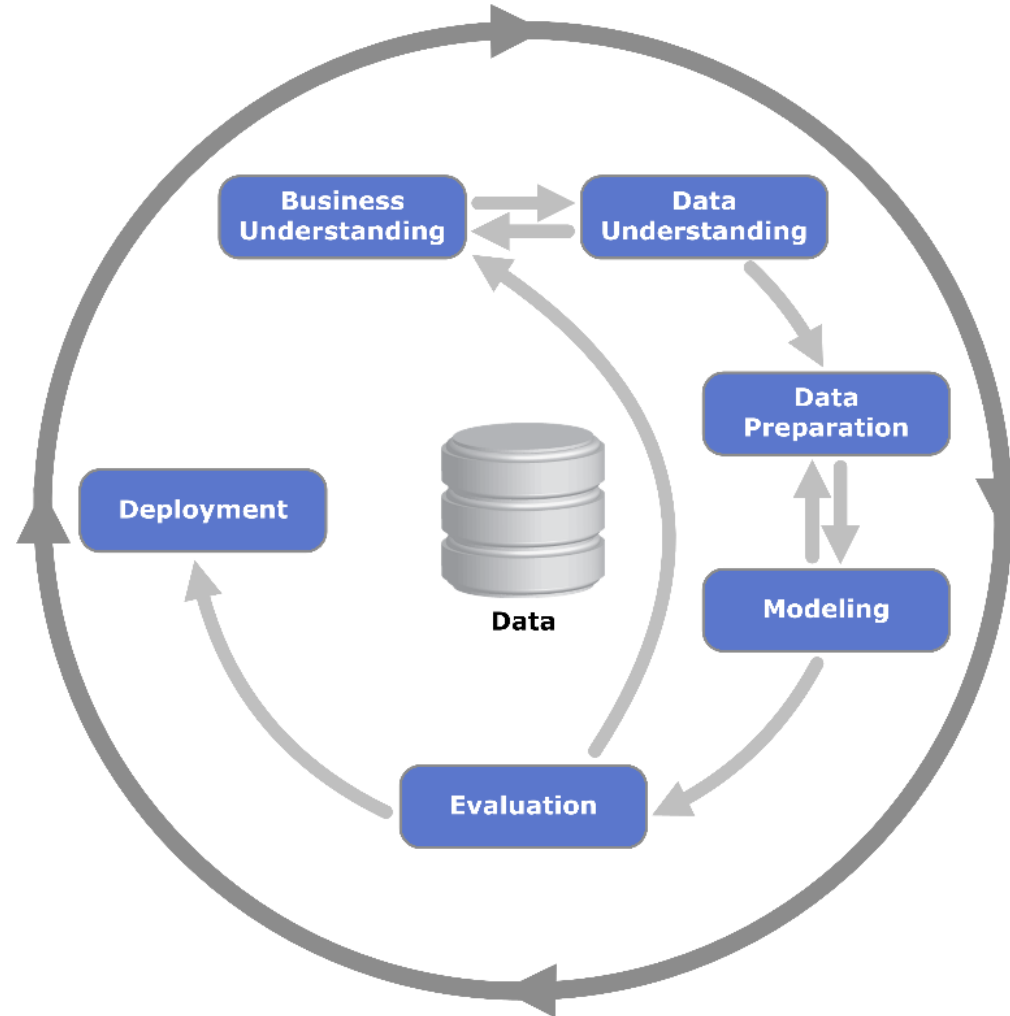
CRISP-DM

SEMMA

KDD

Knowledge Discovery in Databases

Custom



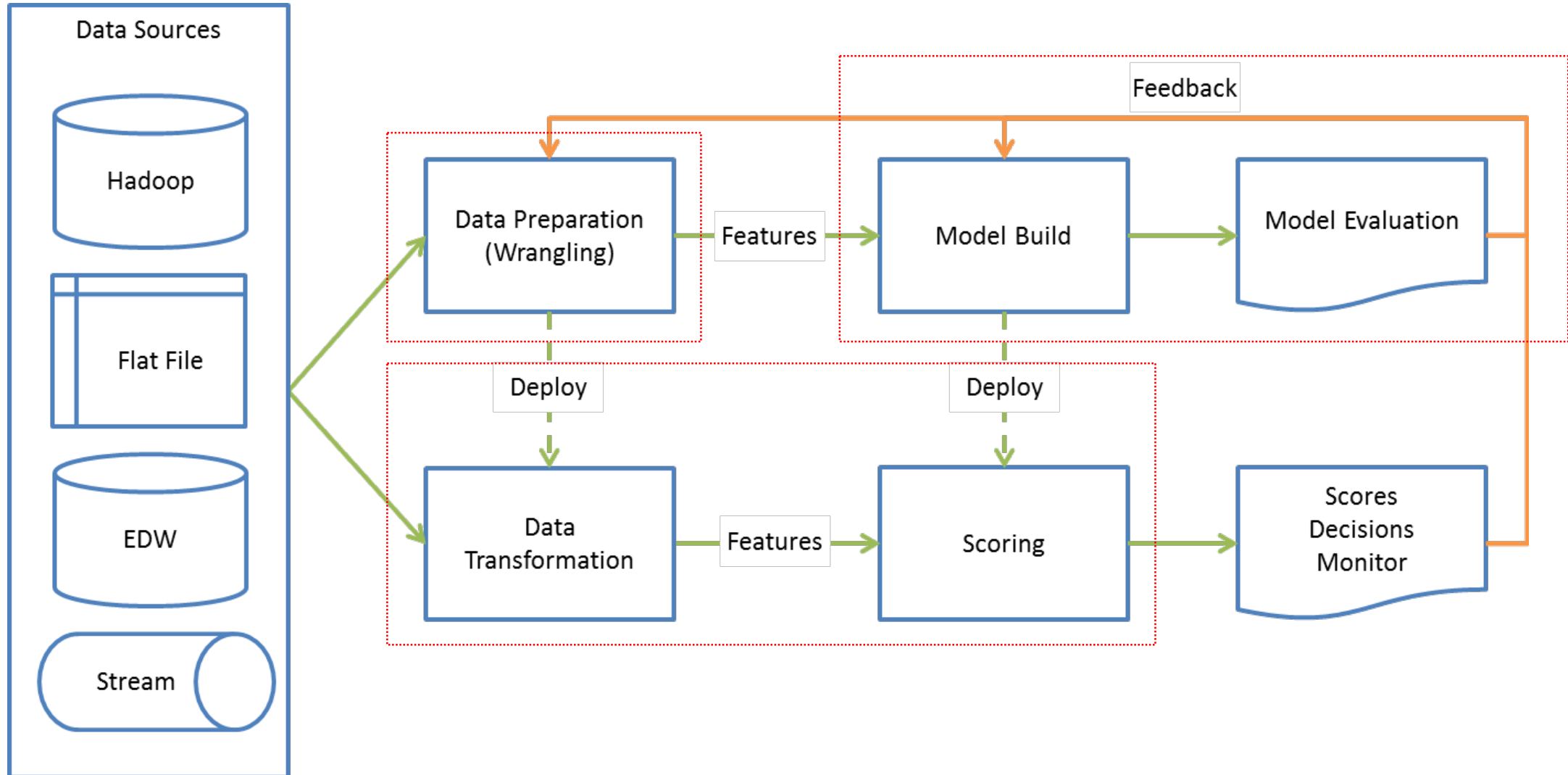
What Is The Business Problem

- Do you want to categorize a record or observation?
- Do you want to predict a numerical quantity?
- Do you want to rank order your records based on some outcome of interest?
- Do you want to identify naturally occurring groupings in the data?

By defining the business problem:

- ✓ Ensure you are collecting the right data
- ✓ Determine what type of model to apply
- ✓ Know when you are done and can go to the next phase

The Data Science Workflow



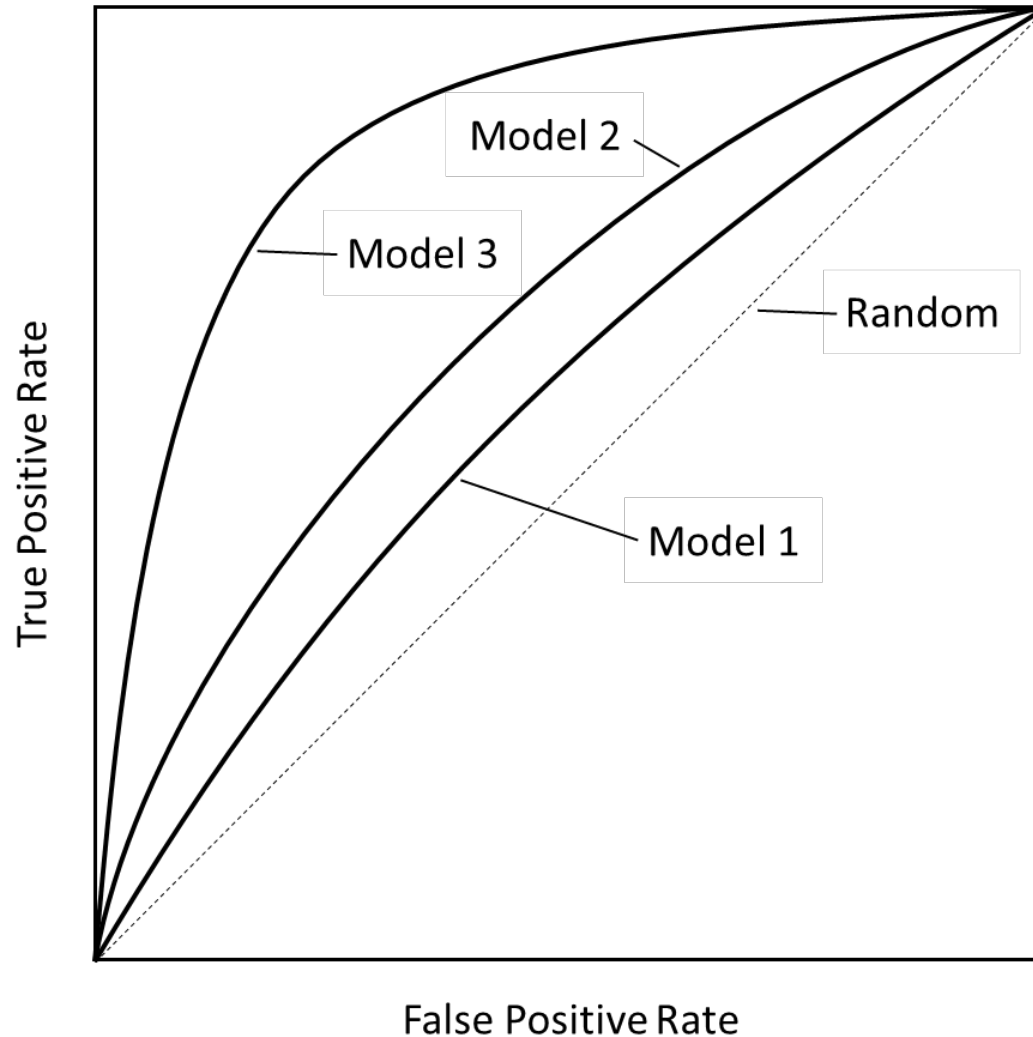
Model Selection

The modeling step is where you apply one of the relevant model types:

- If the data contains known outcomes, then you use a supervised learning algorithm: regression or decision tree (or possibly both).
- If the data does not have known outcomes (or labels), then apply the unsupervised learning algorithm: clustering.

Assessing Model Performance

Interpreting ROC



Which is better:

- Random
- Model 1
- Model 2
- Model 3

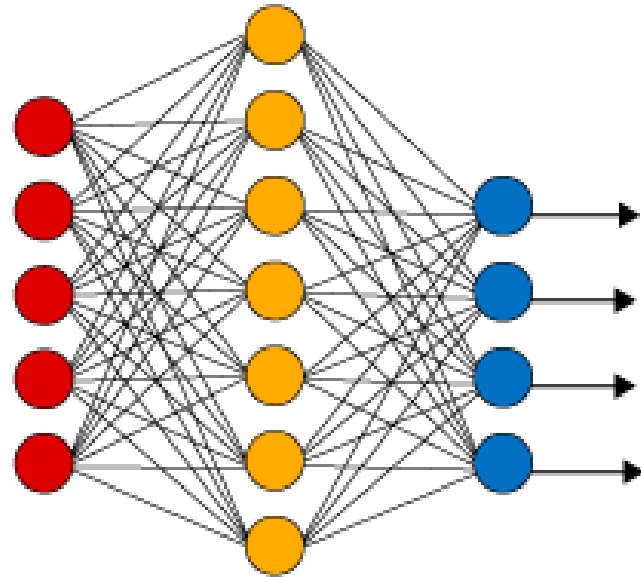


A Sampling of New Model Types

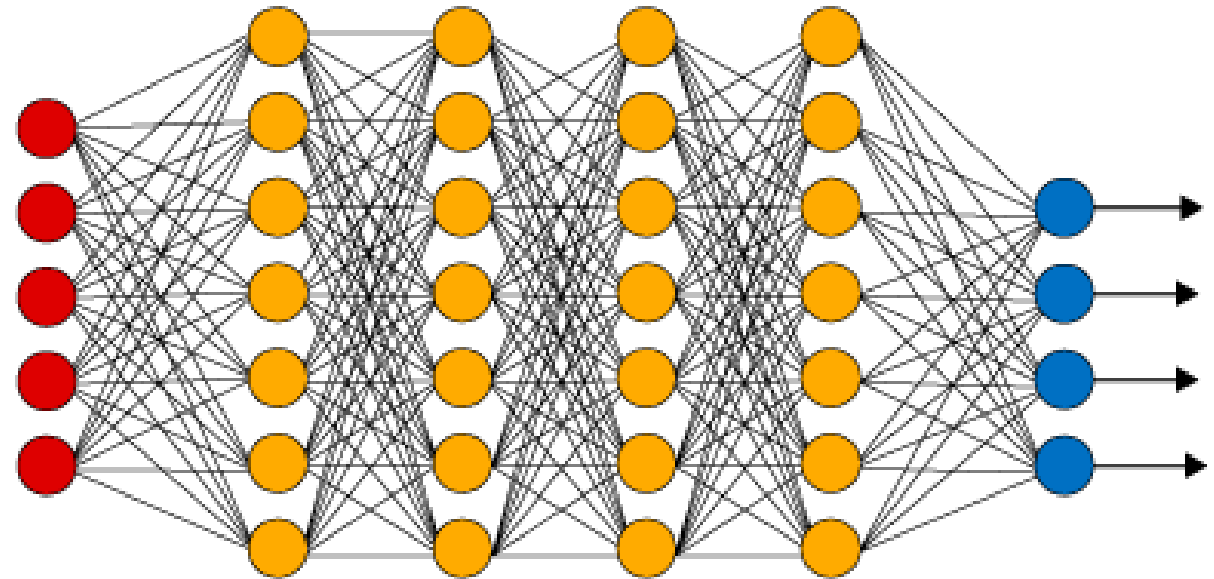
Deep Learning

A New Generation of Artificial Neural Networks

Simple Neural Network



Deep Learning Neural Network



● Input Layer

● Hidden Layer

● Output Layer

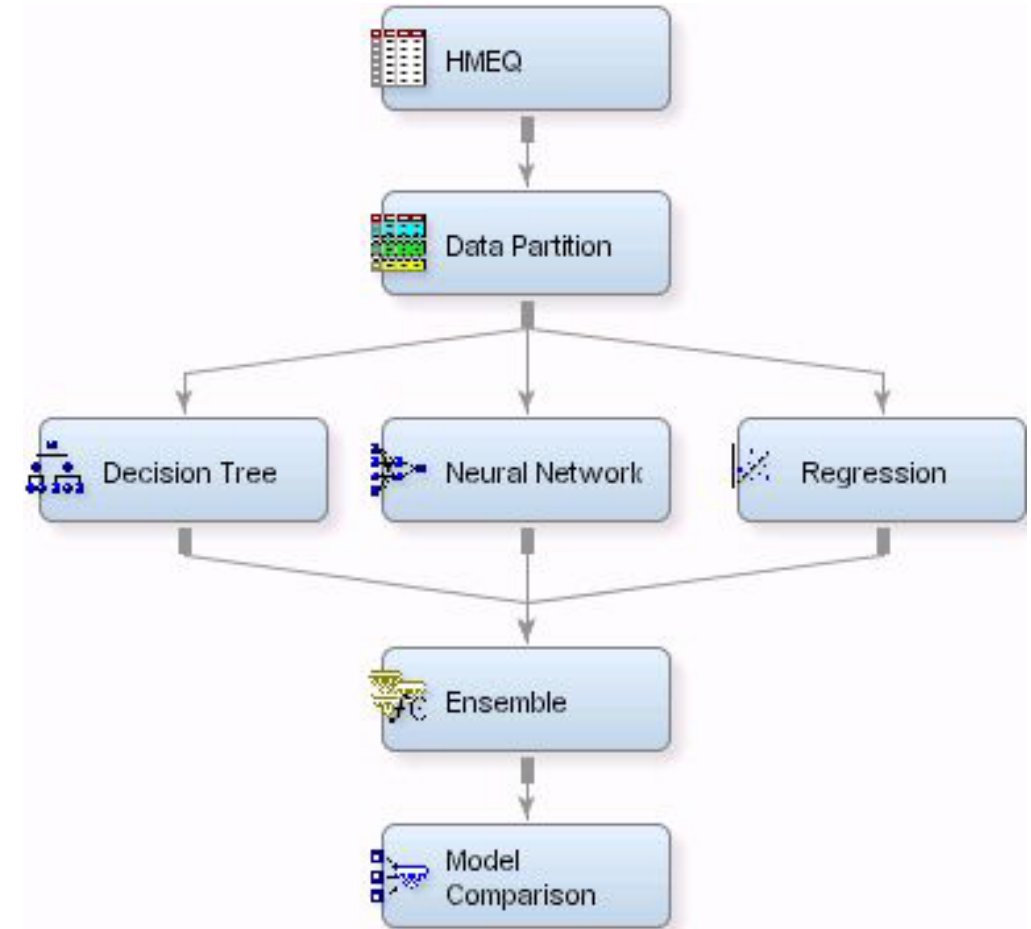
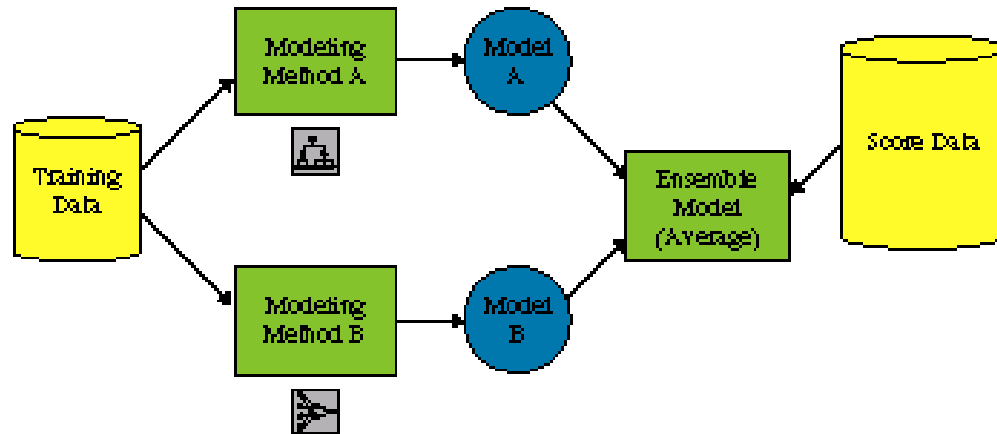
Source: <https://www.quora.com/What-is-the-difference-between-Neural-Networks-and-Deep-Learning>

Deep Learning Use Cases

- Fraud Detection
- Automatic speech recognition
- Image recognition
- Visual art processing
- Natural language processing
- Drug discovery and toxicology
- Customer relationship management
- Recommendation systems
- Bioinformatics
- Mobile advertising
- Image restoration

Ensemble

If one model is good, many models should be better



Source: SAS® Enterprise Miner™ 14.3: Reference Help

Elastic Net

- ✓ High Dimensional Data
- ✓ Correlated Variables



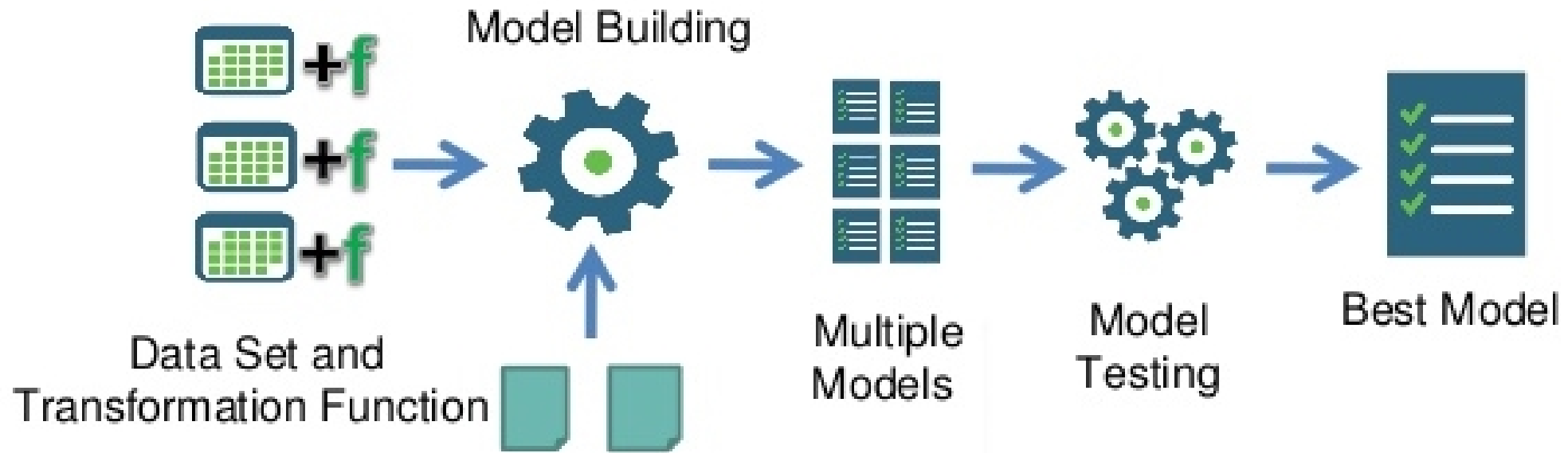
Elastic Net



- ✓ Improve Accuracy
- ✓ Improve Interpretability



Automated Machine Learning



Source: <https://www.slideshare.net/SparkSummit/07-sourabh-chaki>

Where Do We Go From Here

Has the business problem been identified?

- Make sure the focus is on solving the business problem

Which modeling approach to use?

- The core models provide a good starting point and may be sufficient for satisfying the business objective

Are there issues with the data?

- A new model will not fix bad data

Is the model usable/deployable?

- An overly complex model may be difficult to deploy – make sure it's worth it



Q&A

Tom Zougas PhD PEng

tom.zougas@transunion.com

