Data Science & Machine Learning For Quantity Surveyors: A Brief Perspective

Yu Qian Ang, PhD

INTRODUCTION

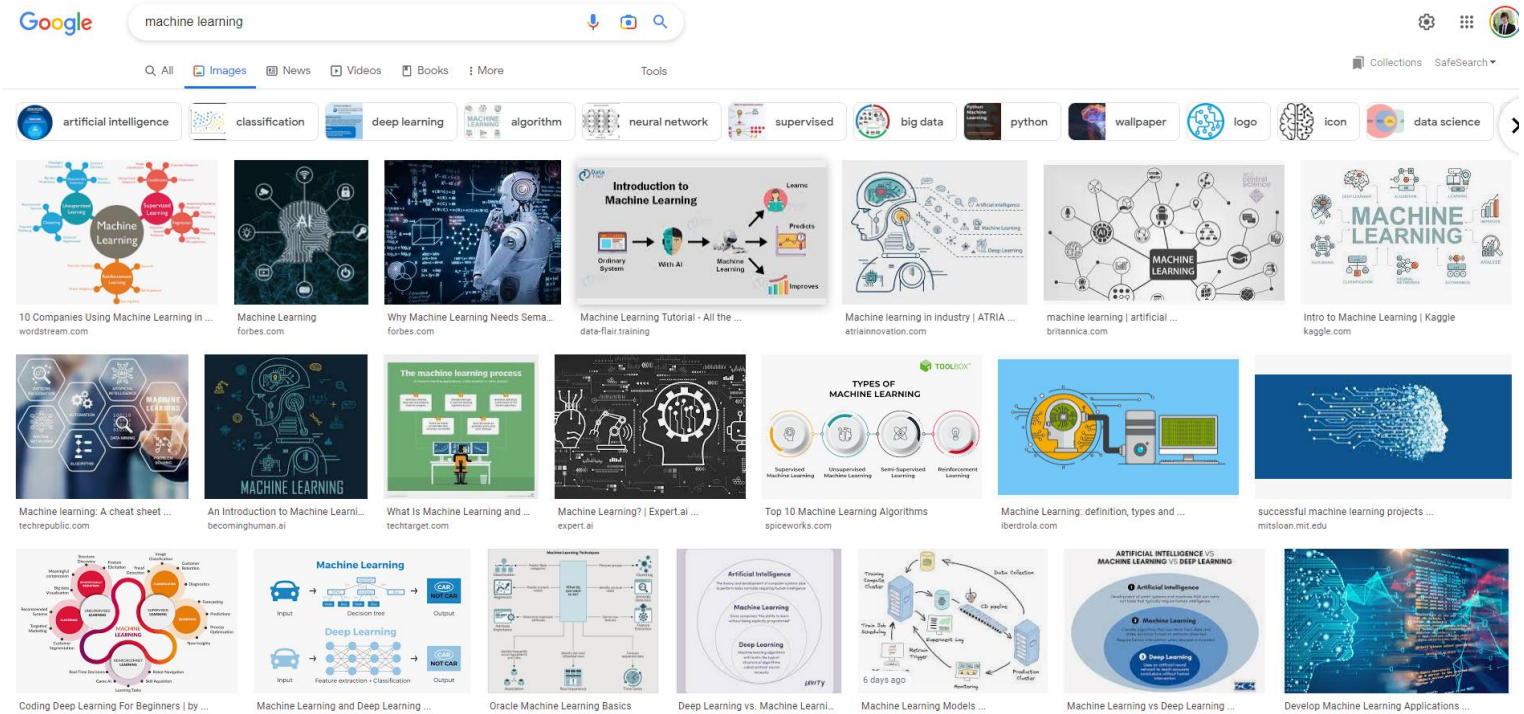
Machine learning as a subfield of artificial intelligence

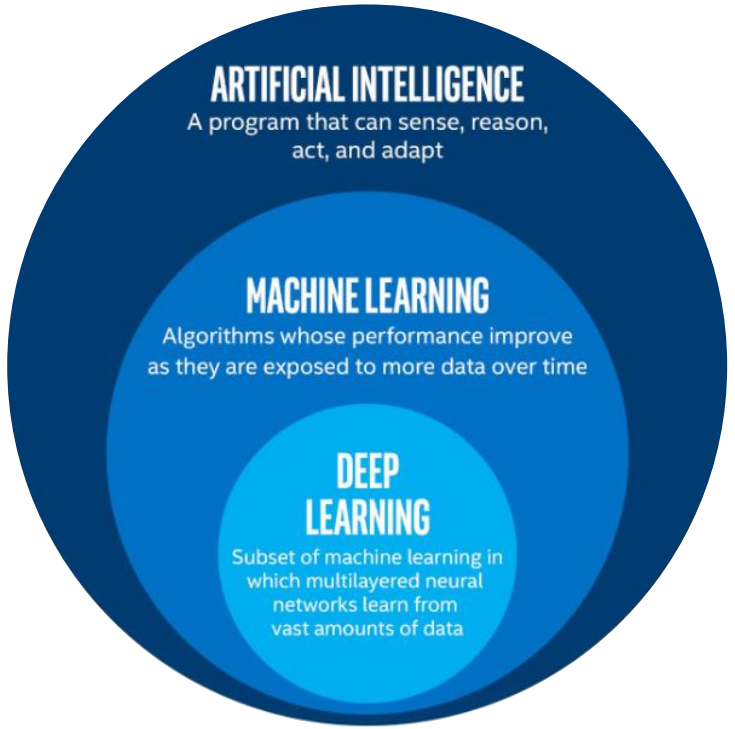
Everyone talks about it

Nobody really knows how to do it

Everyone thinks everyone else is doing it

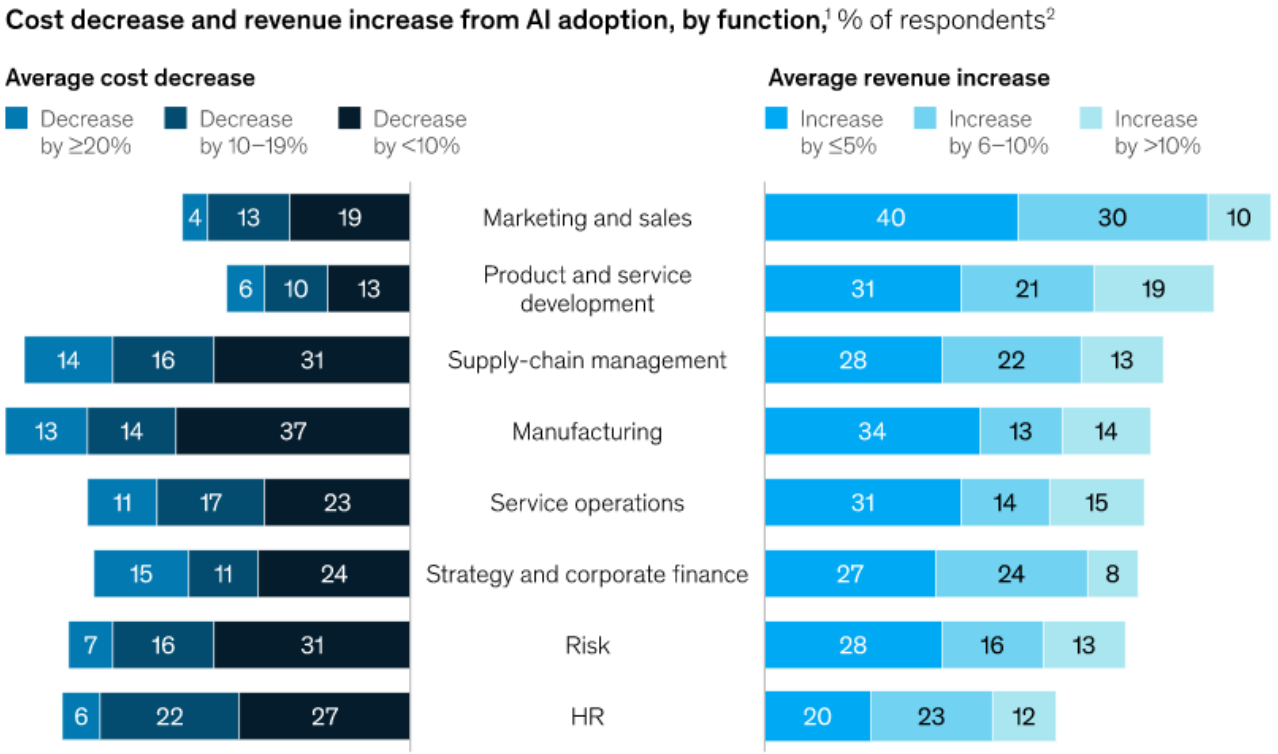
So everyone claims they are doing it too



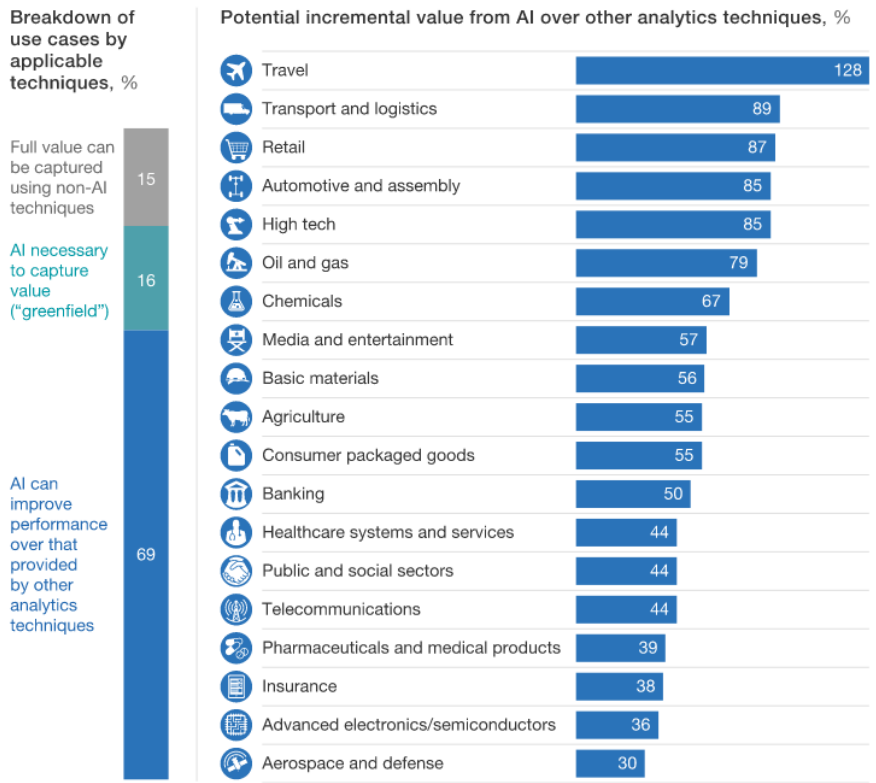


INTRODUCTION

Organizations report lower cost and higher revenue



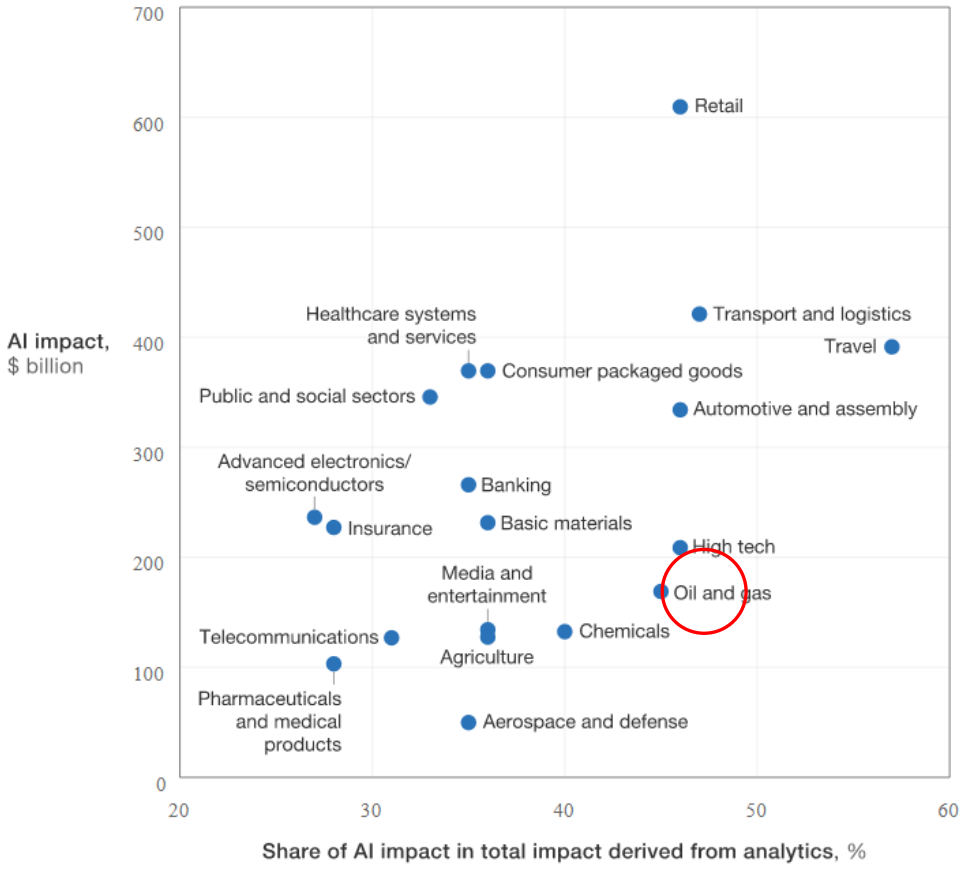
Value based on sectors



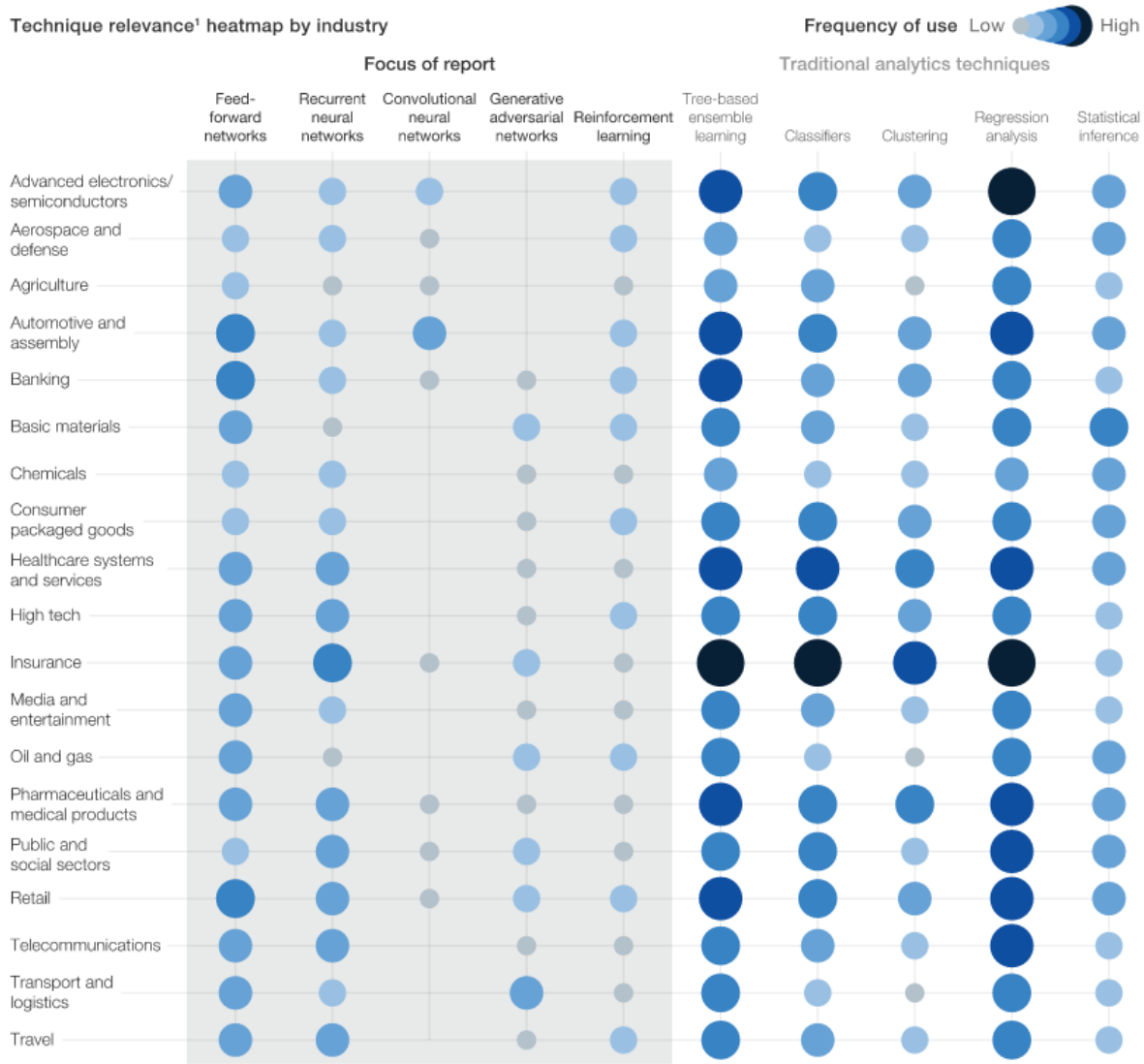
Charts from McKinsey Global Institute

FRAMEWORK

Impact based on sectors



Technique relevance based on sectors



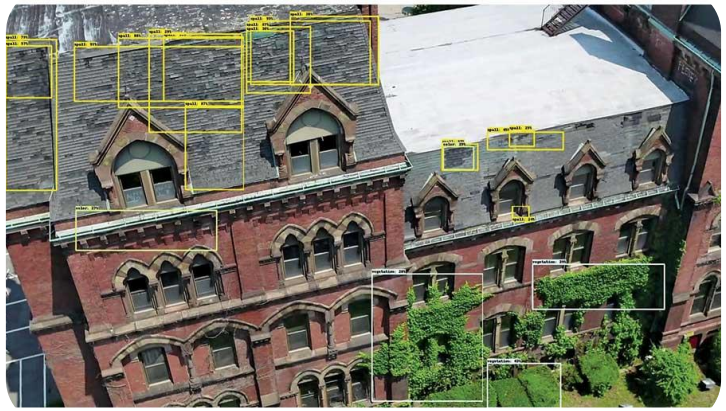
Charts from McKinsey Global Institute

INTRODUCTION

Machine learning already used in many Architecture, Engineering and Construction (AEC) applications

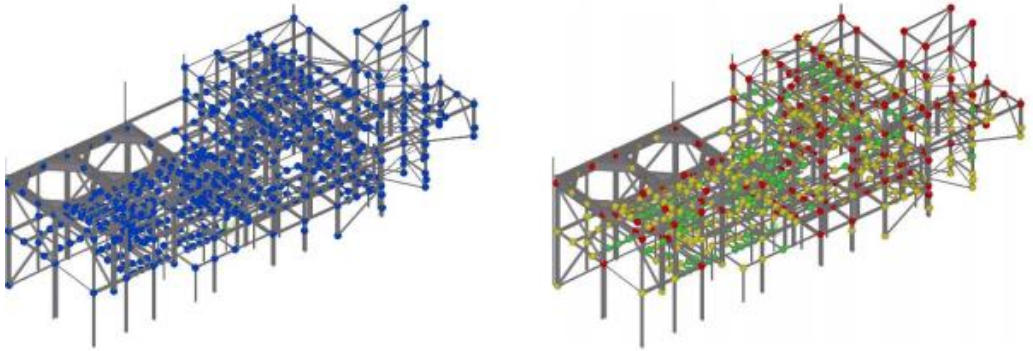
1 Façade Inspection

Rebenstone, 2020



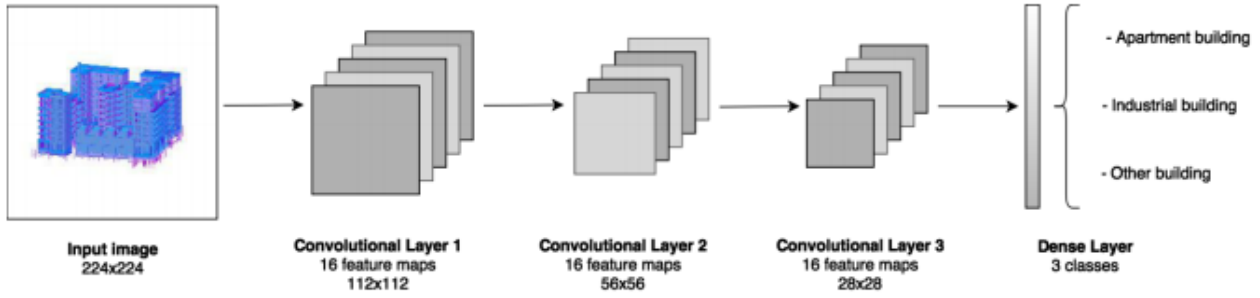
2 Generating Steel Connections

Lomio et al., 2018



3 Classification of BIM Structures

Helminen et al., 2018



FRAMEWORK

Supervised Learning	
Regression **Outcome: Continuous (example: predicting cost)**	
Linear Regression	Basic fast, linear model
Bayesian Linear Regression	Linear model, small dataset
Decision Tree Regression	Fast, able to include both categorical and numerical predictors
Tree-based Ensemble Models (e.g. Random Forest Regressor)	Fast, accurate Able to include both categorical and numerical predictors
Boosted Ensemble Models (e.g. XGBoost, LightGBM)	Accurate, longer training time than typical decision-tree models
Neural Network Regressors	Requires larger training dataset but can be potentially accurate

Supervised Learning	
Classification **Outcome: Discrete (example: classify type of variation order)**	
Logistic Regression	Basic, fast
Decision Tree Classifier	Fast, can include both categorical and numerical predictors
Tree-based Ensemble Models (e.g. Random Forest Classifier)	Fast, accurate Able to include both categorical and numerical predictors
Boosted Ensemble Models (e.g. XGBoost, LightGBM)	Accurate, longer training time than typical decision-tree models
Neural Network Classifiers	Requires larger training dataset but can be potentially accurate

Unsupervised Learning	
Clustering **Outcome: Clusters (example: group similar cost items)**	
K-Means	Basic, fast. Need to select K
Hierarchical Clustering	Merge data points until desired number of clusters reached
DBSCAN	Find high-density regions
Mean Shift	Find and adapt centroids
Gaussian Mixture Models	Summarizes a multivariate probability density function with a mix of Gaussian probability distributions

predict values

predict categories/groups

discover patterns/structures

Method Type / Purpose

reduce data dimensions

analyse text

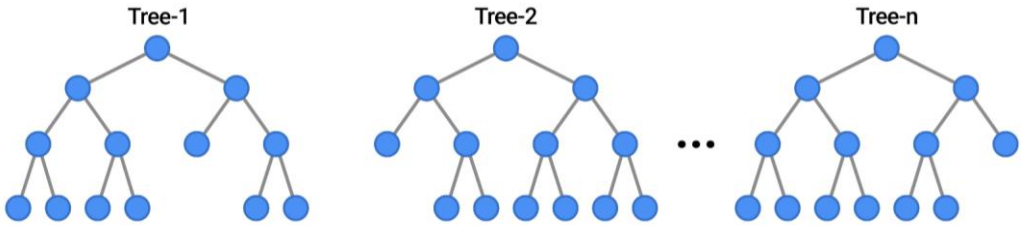
Dimensionality Reduction	
Outcome: Reduce Dimension (example: Reduce predictors)	
Principal Component Analysis	Basic, popular, fast
t-SNE	Nonlinear dimensionality reduction, stochastic
Singular Value Decomposition	Fast, works for sparse data
Linear Discriminant Analysis	Multi-class classification algorithm that can be used for dimensionality reduction

Text Analytics / Natural Language Processing	
Outcome: Derive Information From Text (example: analyse contract documents)	
Latent Dirichlet Allocation	Unsupervised topic modelling
Word2Vec	Convert words to values for use in NLP tasks
Transformers	Deep learning model with self-attention, differentiate weighting

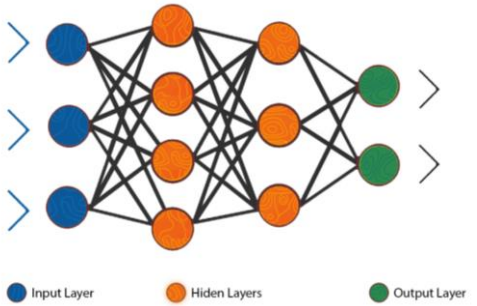
SUPERVISED

Supervised learning categorized by the use of labelled training datasets to train algorithms

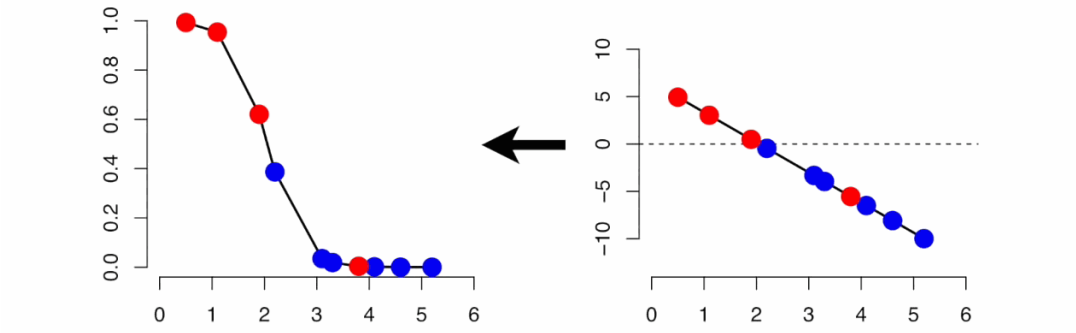
Tree-based Models



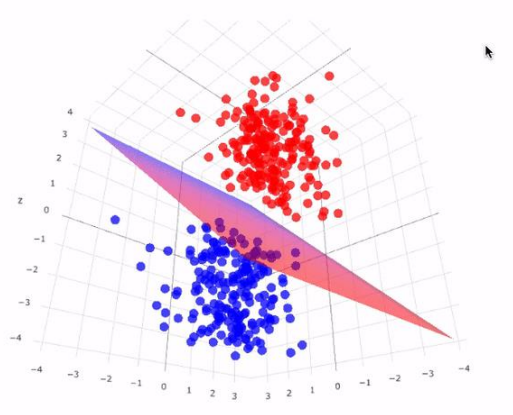
Neural Networks



Logistics Regression



SVM Classifier



Images from: Tensorflow blog, Analytics Vidhya, Ashwin Raj, KD Nuggets

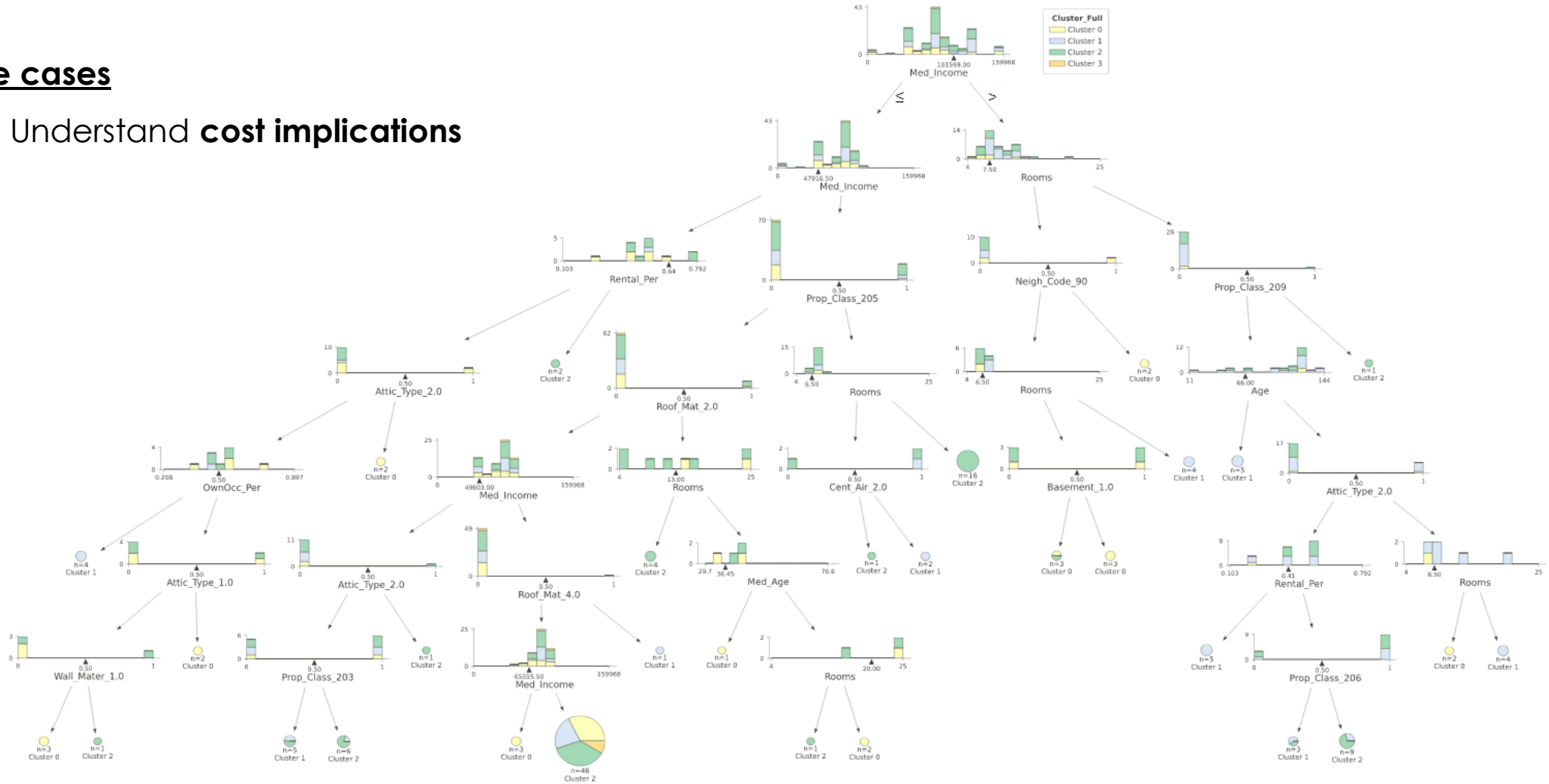
SUPERVISED

Tree-based ensemble methods

Regression Trees (Random Forest)

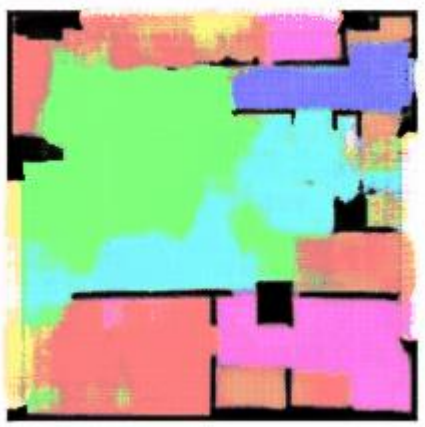
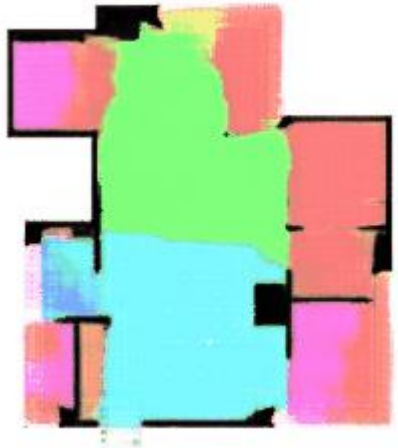
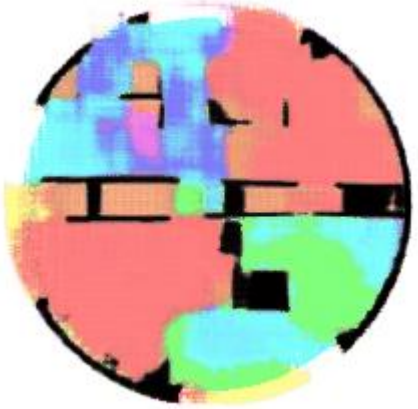
Use cases

1. Understand **cost implications**

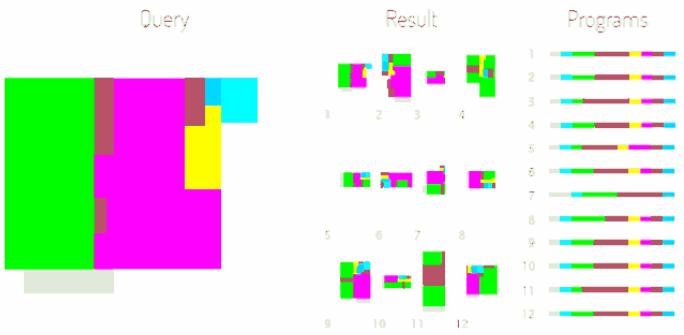


SUPERVISED

Generative Adversarial Networks for Floor Plans – by Stanislas Chaillou (Harvard GSD)


Generating floor plan



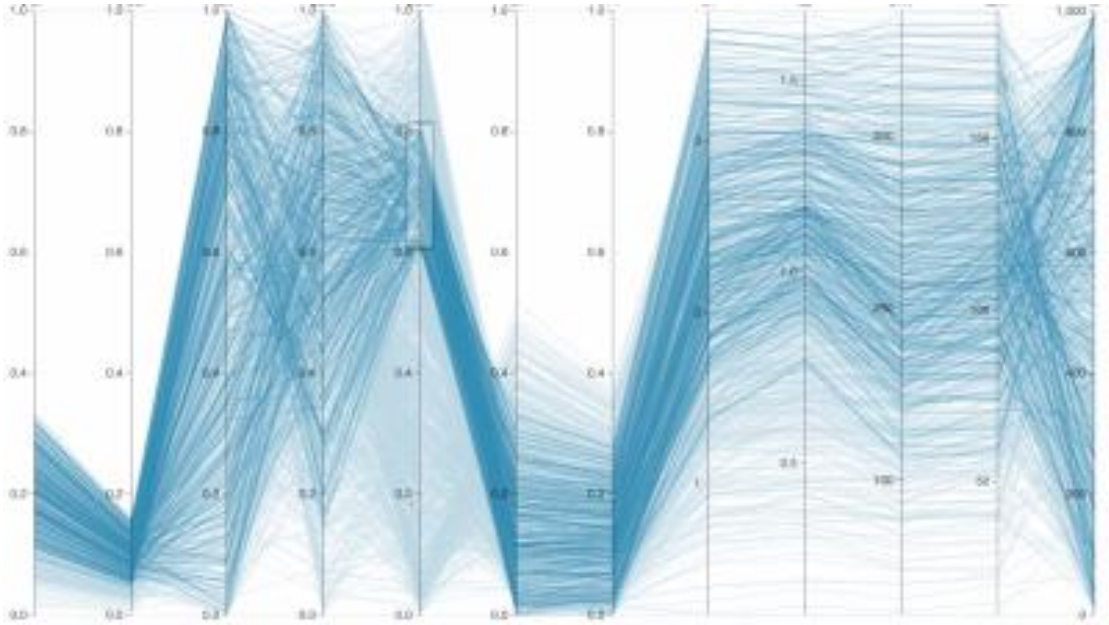
SUPERVISED

Generative Adversarial Networks for entire buildings

Generating entire buildings + cost plans

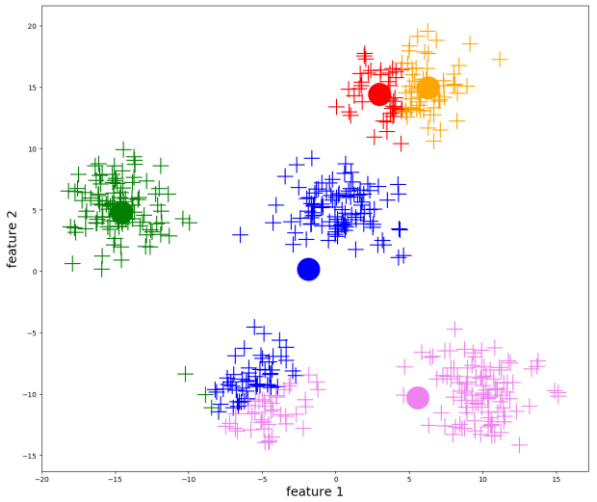
Use cases

1. Automated, **intelligent cost plans**
2. Automatically **re-generate / repair elements**
3. Recommendations and **insights to support decisions**

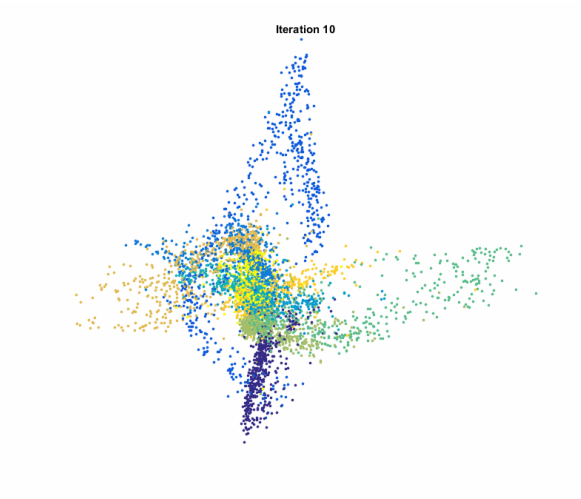


UNSUPERVISED

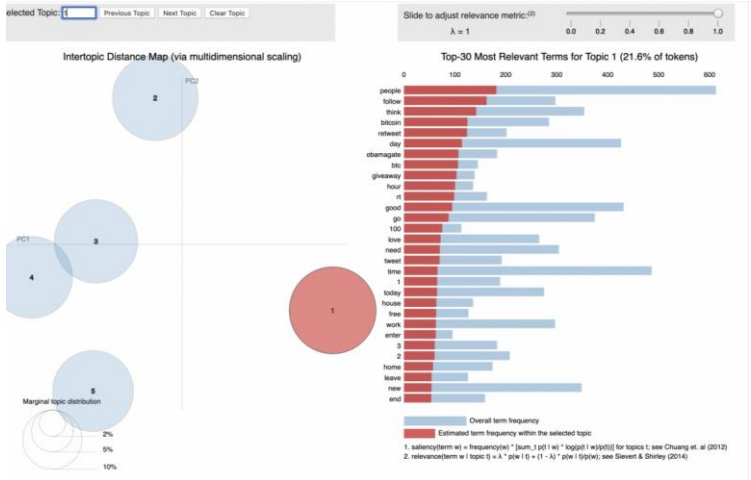
Clustering



TSNE



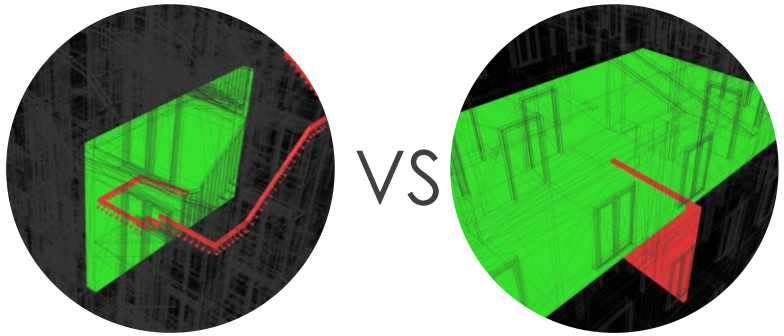
Topic Modeling

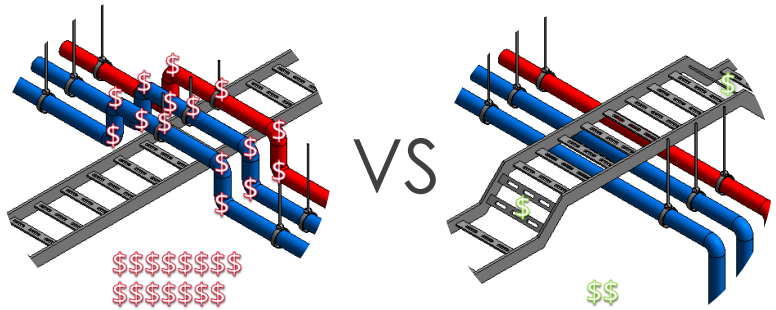


Images from: Pritee Dharme, Yinsen Miao, Amin Azad

UNSUPERVISED

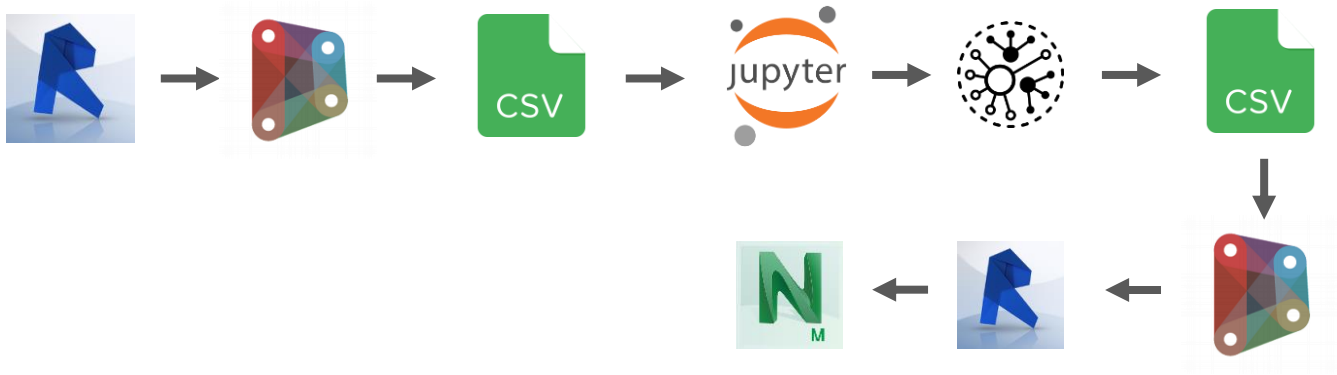
Some conflicts involve **higher cost & time implications**



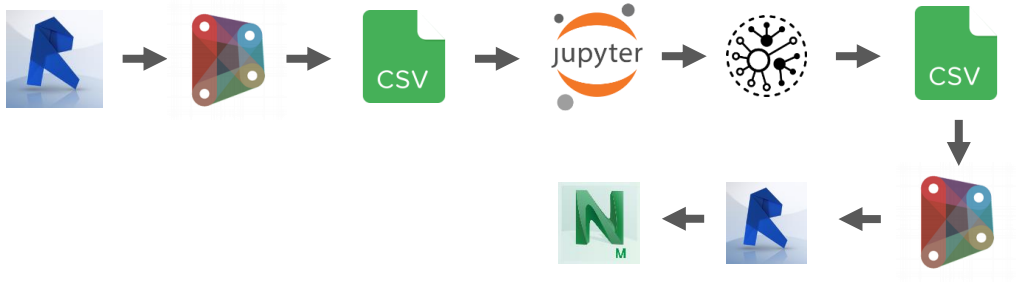


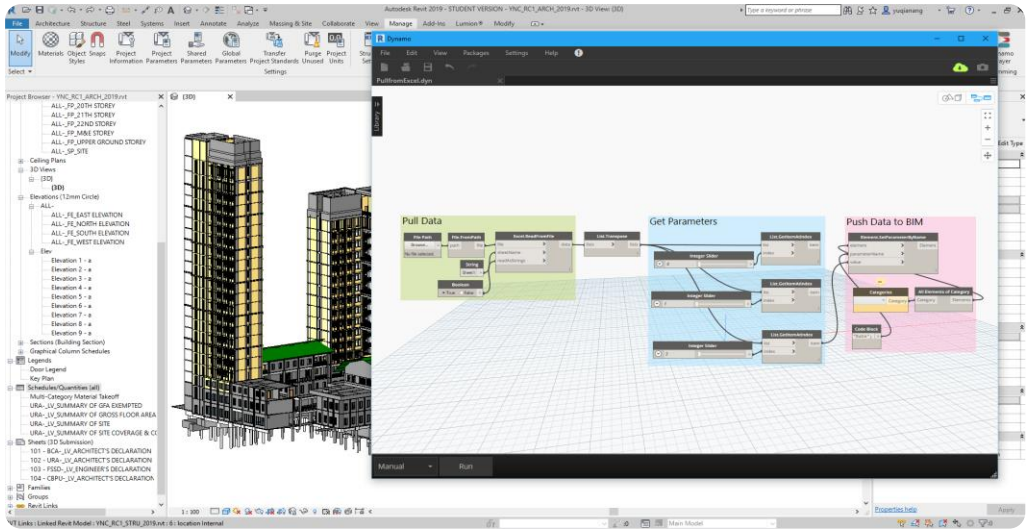
(image source: lodplanner)

Can we design a process where algorithms will identify these?

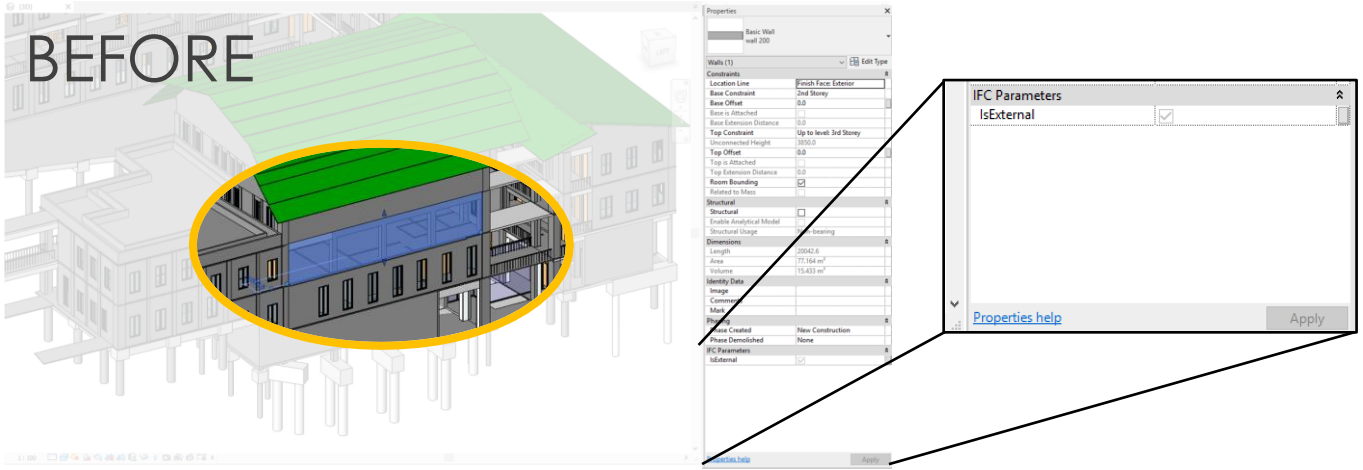


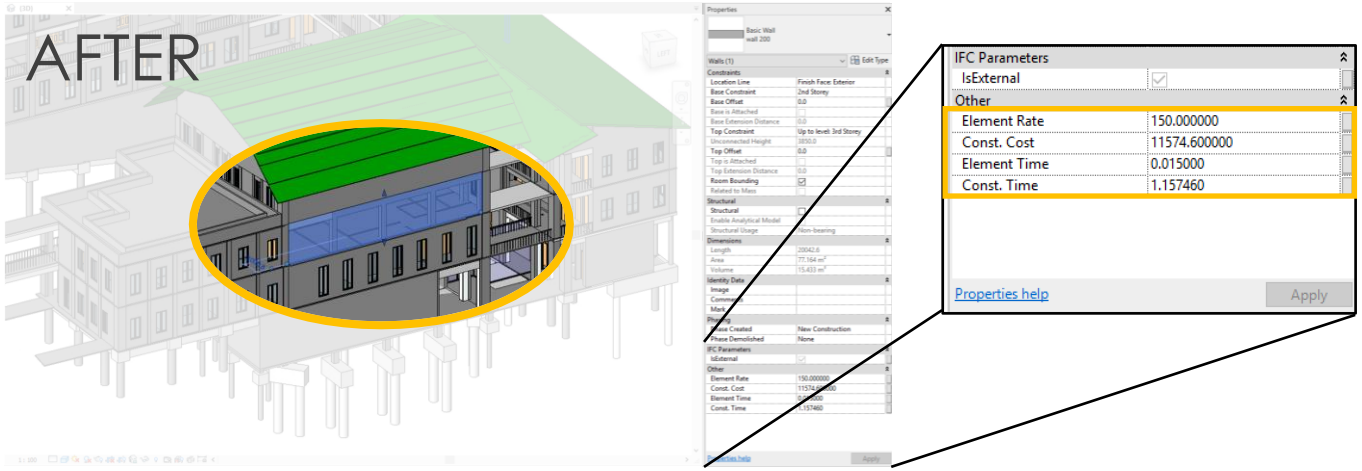
UNSUPERVISED





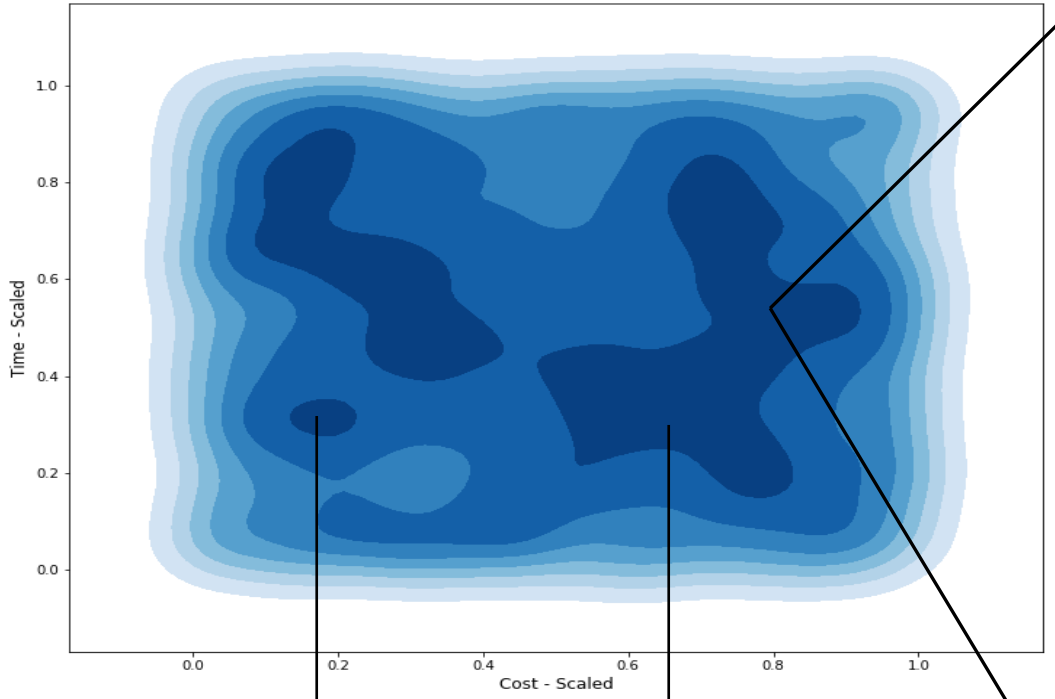
Input **cost (rates)** and **time** parameters into BIM

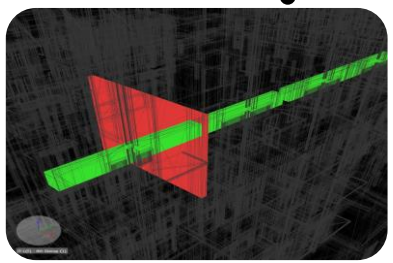


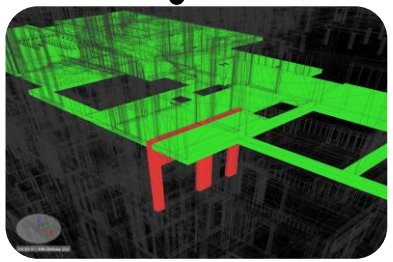


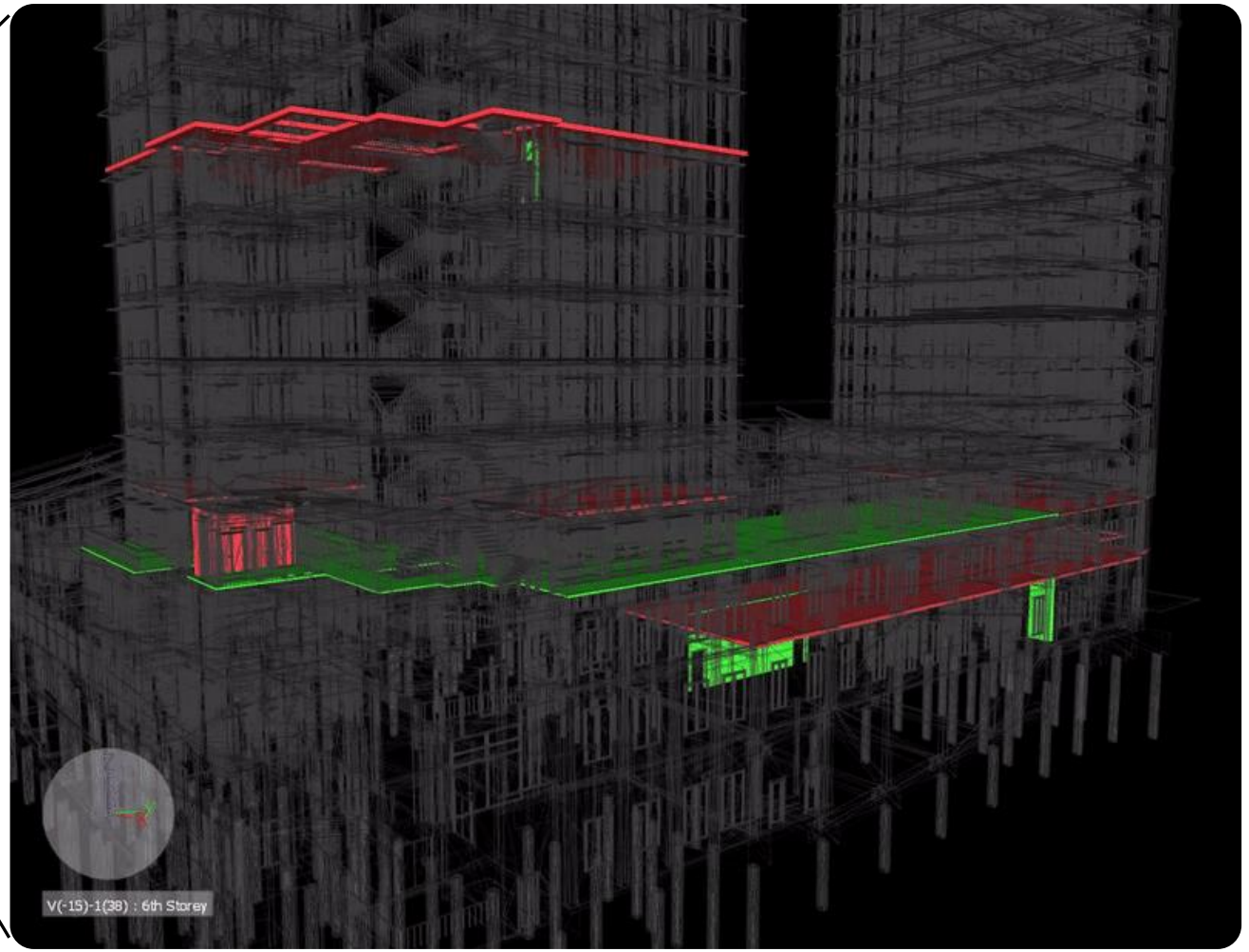
UNSUPERVISED

Run clustering algorithm – cluster based on **cost & time**



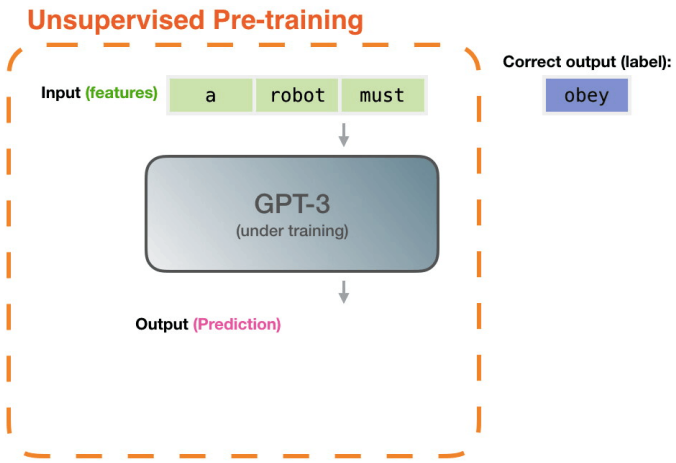






OTHERS

Training language models (GPT-3)



Use cases for Natural Language Models

1. Highly intelligent **auto-complete for tender documents**
2. Automatic **generating contracts** based on inputs

```
import requests
r = requests.post(
    "https://api.deepai.org/api/text-generator",
    data={
        'text': 'The quantity surveyor for the construction project made a mistake in his cost estimates.',
    },
    headers={'api-key': 'quickstart-QUdJIGlzIGNvbWluZy4uLi4K'}
)
```

```
output = r.json()
print(output['output'])
```

```
The quantity surveyor for the construction project made a mistake in his cost estimates. The first round of costs to date, for an initial five years worth $25 million, had been the result of two "lowest estimate of any of the contracts," he explained.

As costs were put in place, though, "it was very difficult for me to make sense of the new contract for a contractor." This gave the project time to build up its work.

"By doing the lower estimate, we were able to avoid the uncertainty we had over when to start using a subcontractor's contract and how short it would have to be to start working fulltime," he explained.

This helped the contractors avoid the same financial troubles that prevented them from hiring a full-time engineer at the facility. This was a factor that led to the design of the facility, especially at a time when the state was trying to build its own electricity grid.

He also noted that no matter which contractor was hired, there was often not enough funding available to get the facility up to speed. This could be a factor in the lack of funds for a large project like the solar panel plant.

Budget

In the end, B.C.'s largest project, BC West Regional Transmission, raised its funding goal at $7.6 million.

The total cost for this program, which was announced in December 2011, raised $2 million. One year earlier, B
```

GPT-3 animation from Jay Alammar

PROBLEMS

Challenges + Opportunities

CHALLENGES X **OPPORTUNITIES**

Project-based nature of construction projects

Culture of **valuing seniority and experience** over empirics

Reluctance to accept new technology

Profession takes **very long to mature**

Reliance on experience rather than data

Challenges X Opportunities

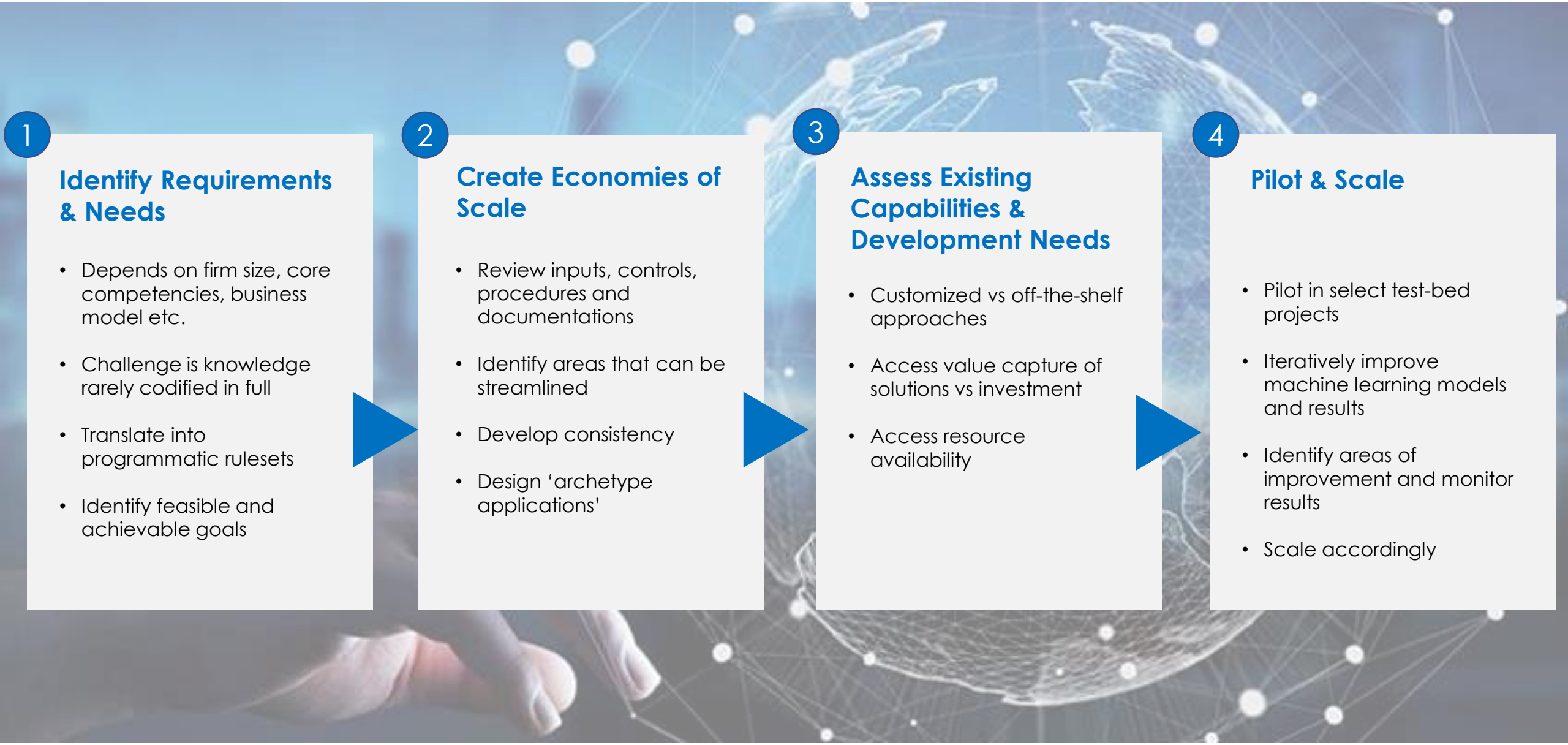
A lot of **manual work** (quantity take-offs etc)

Projects are **data rich**

Projects are **data rich**

Implementation for value capture

IMPLEMENTATION



1

Identify Requirements & Needs

- Depends on firm size, core competencies, business model etc.
- Challenge is knowledge rarely codified in full
- Translate into programmatic rulesets
- Identify feasible and achievable goals

2

Create Economies of Scale

- Review inputs, controls, procedures and documentations
- Identify areas that can be streamlined
- Develop consistency
- Design 'archetype applications'

3

Assess Existing Capabilities & Development Needs

- Customized vs off-the-shelf approaches
- Access value capture of solutions vs investment
- Access resource availability

4

Pilot & Scale

- Pilot in select test-bed projects
- Iteratively improve machine learning models and results
- Identify areas of improvement and monitor results
- Scale accordingly

CONCLUSION

Takeaways

Not Software
Common software should be able to interface with machine learning code

Importance of Technical Skillsets + Domain Knowledge
Programming literacy important in modern day AEC domains

Value-added Services
Consultancy hyper-competitive, important to drive higher value capture

Competitive Advantage
QS firm and software vendor that figure out how to incorporate these workflows poised to be market leader


Thank You