Data science:
How to leverage data for superior decision making

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

100
1918 · 2018

Jacomine Grobler
jacominegrobler@sun.ac.za


DATA

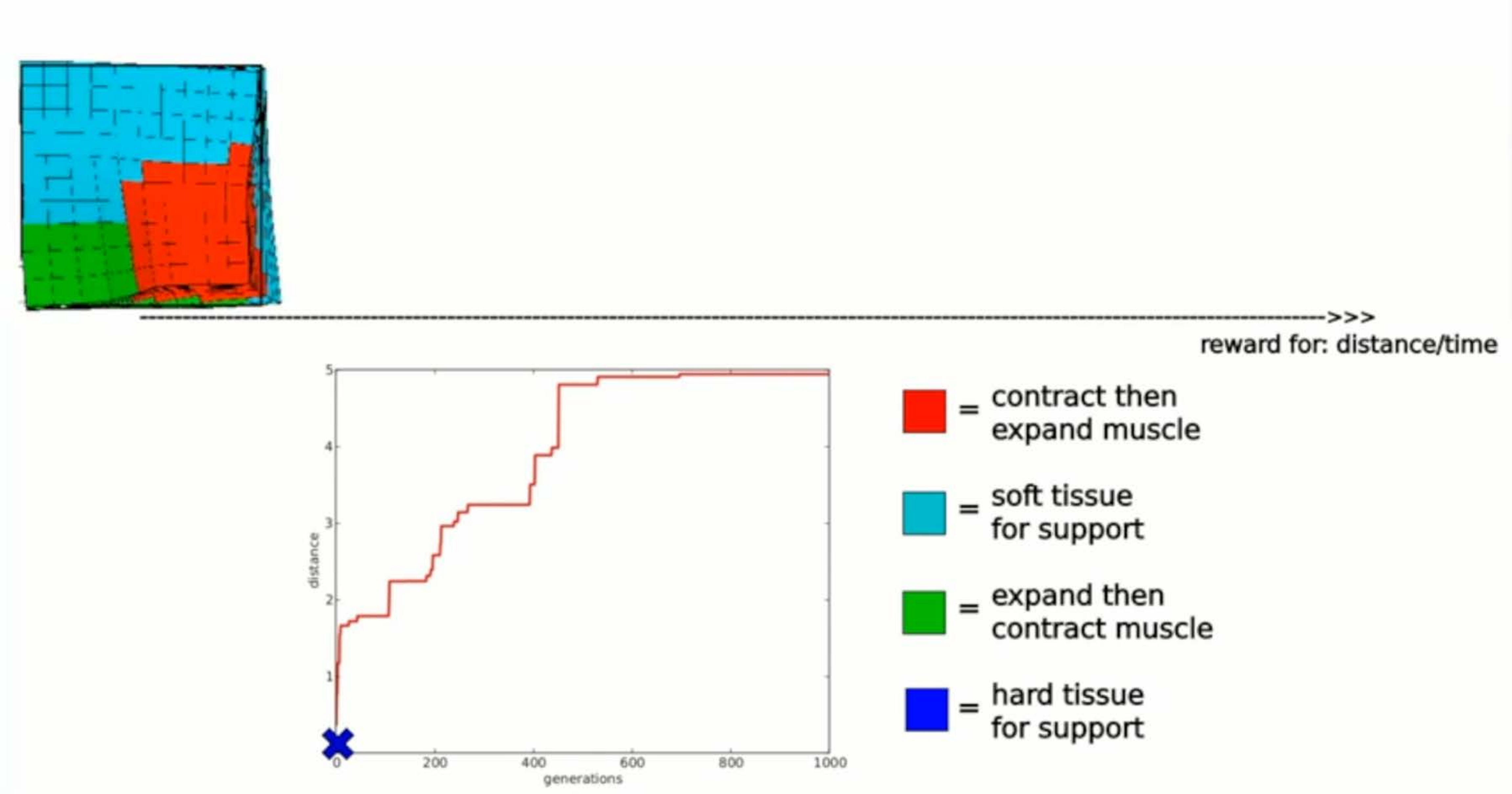
Artificial Intelligence

Deep Blue

Garry Kasparov

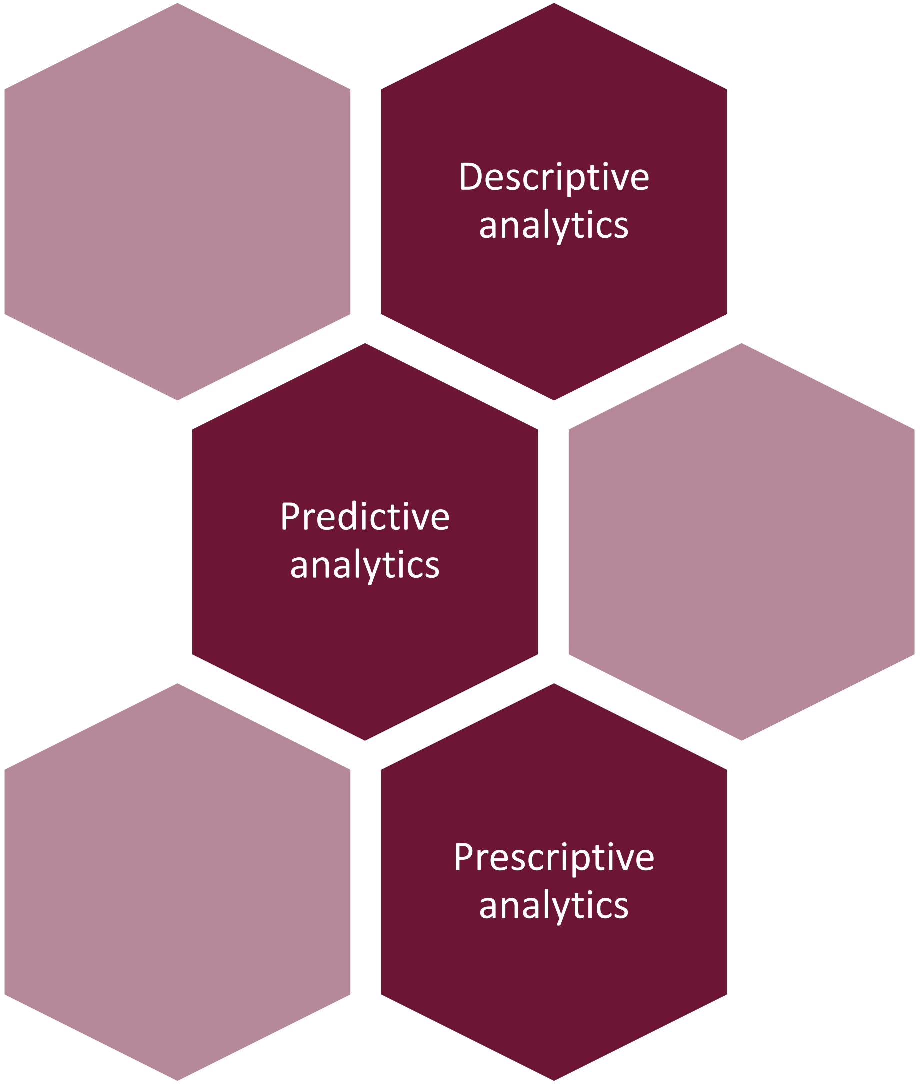
AlphaGo

EINRIDE



Acknowledgements: Cheney, Nick, Robert MacCurdy, Jeff Clune, and Hod Lipson. "Unshackling evolution: evolving soft robots with multiple materials and a powerful generative encoding." *ACM SIGEVOlution* 7, no. 1 (2014): 11-23.

Data Science

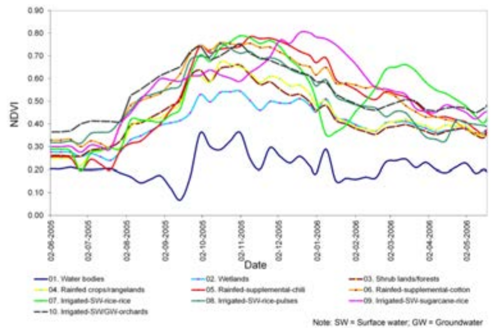


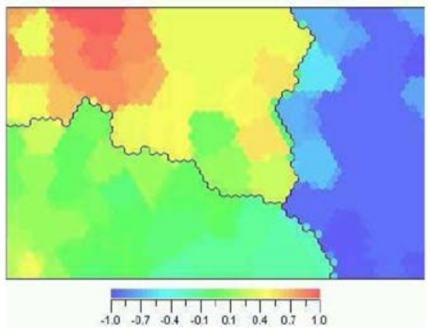
UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY
100
1918·2018

Acknowledgements: Yolandi Le Roux; University of Pretoria

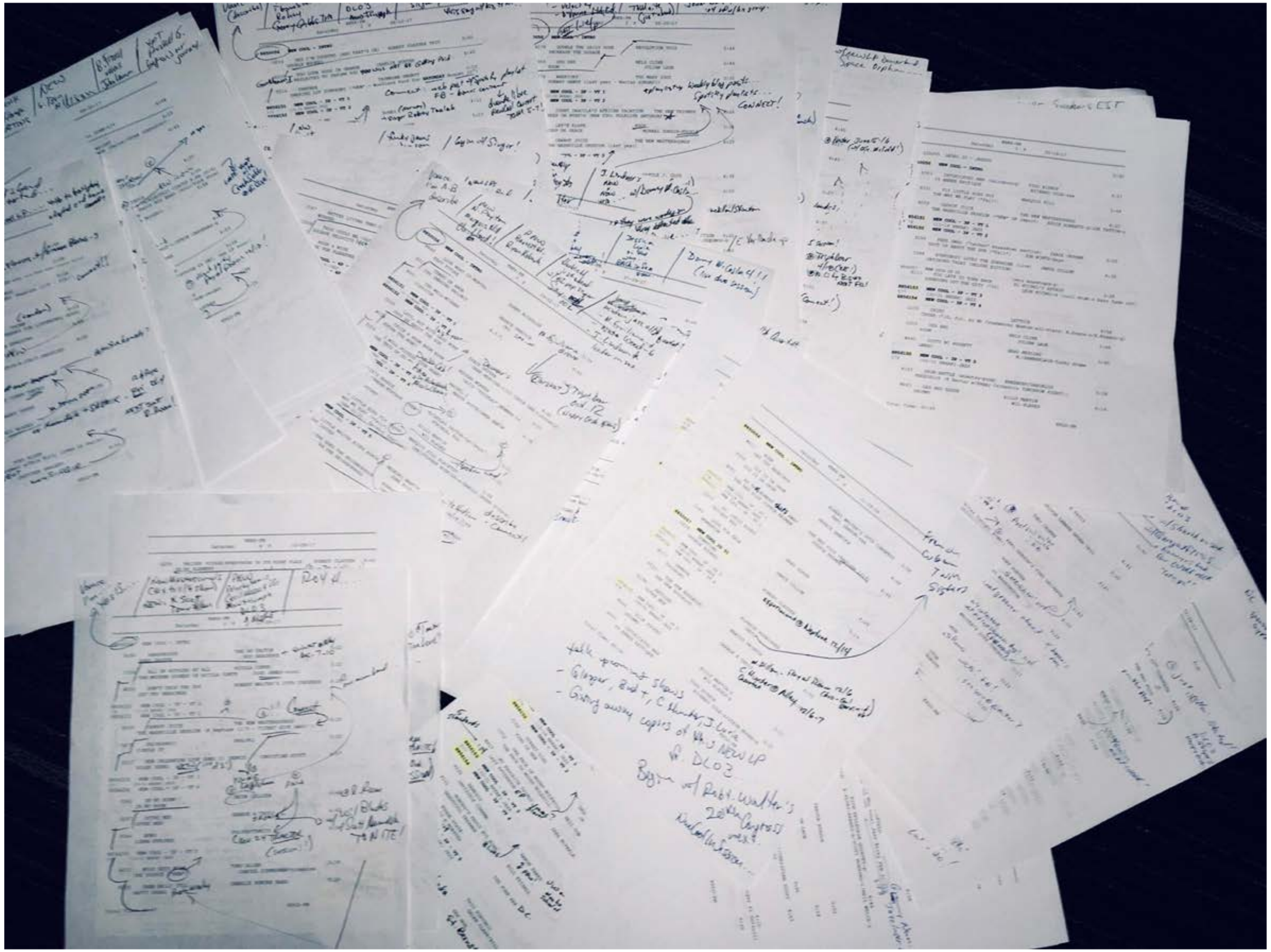
A Predictive Model from the Agricultural Industry

- Plant type
- Water
- Fertilizer
- Location
- History
- Soil type



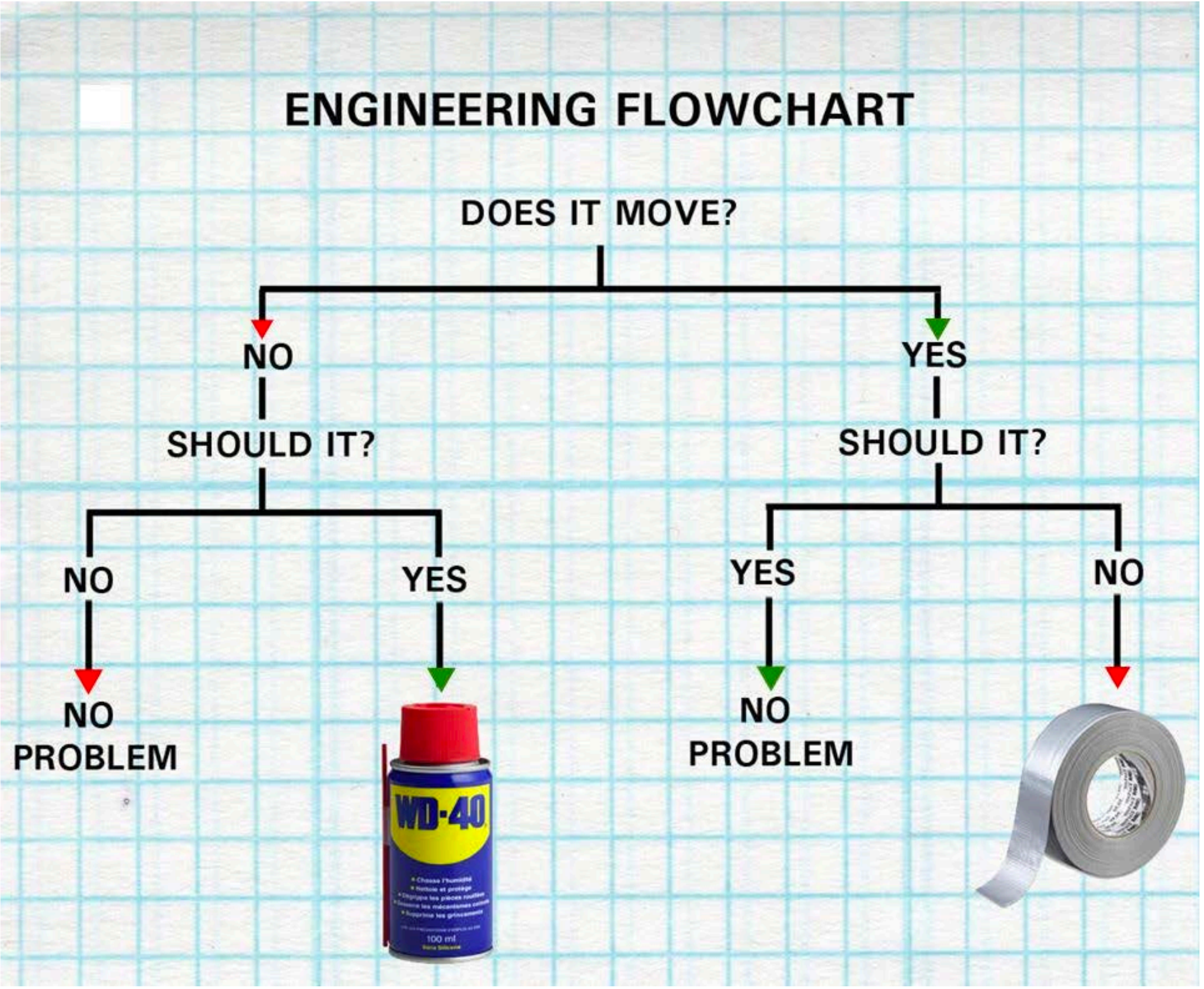


Acknowledgements: Yolandi Le Roux; University of Pretoria

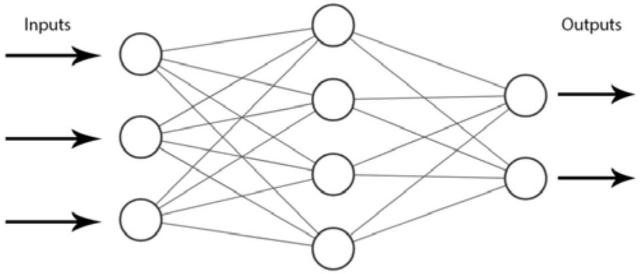


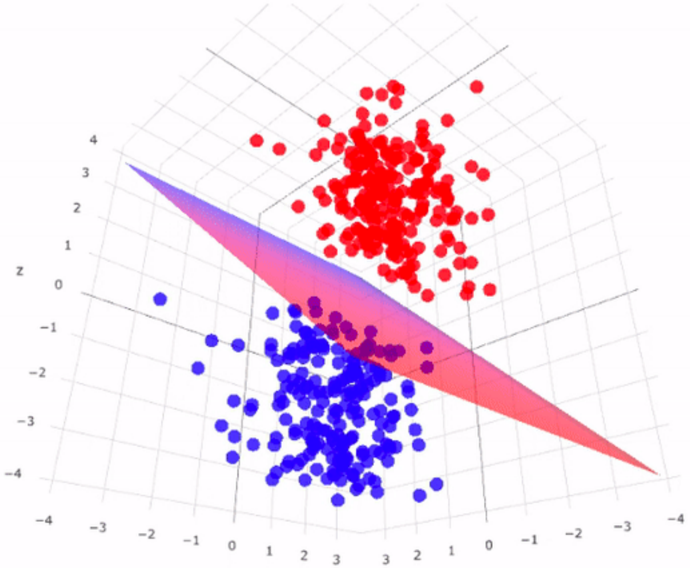
Acknowledgements: Yolandi Le Roux; University of Pretoria

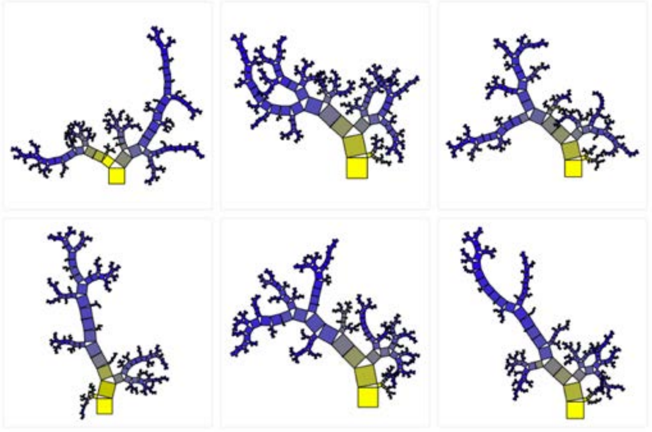
Decision Trees



More Algorithms







Results

Classifier Type	Algorithm	Accuracy
Bayesian	Bayes Net	40.89%
	Naïve Bayes	36.89%
	Naïve Bayes Updateable	36.89%
Functions	Logistic	35.15%
	Multi layer Perception	55.92%
	Simple Logistic	33.79%
	SMO	38.34%
Lazy	IBK	93.90%
	Star	90.16%
	LWL	30.51%
Meta	Attribute Selected Classifier	81.79%
	Bagging	73.59%
	Classification Via Regression	74.59%
	Filtered Classifier	61.11%
	Iterative Classifier Optimiser	44.72%
	Logit Boost	44.72%
	Multi Class Classifier	34.15%
	Random Committee	94.17%
	Randomizable Filtered Classifier	94.17%
	Random Sub Space	68.49%
Rules	Decision Table	63.48%
	JRip	68.76%
	Part	88.25%
Trees	Hoeffding Tree	36.16%
	J48	87.70%
	LMT	88.98%
	Random Forest	94.17%
	Random Tree	93.90%
	REP Tree	63.39%

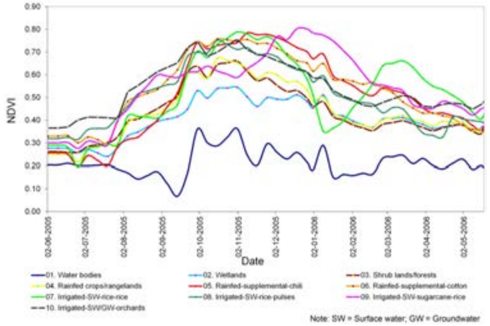
UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY
100
1918·2018

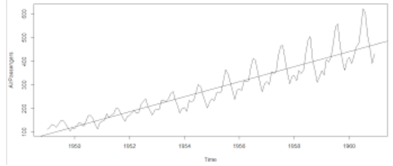
94% accuracy obtained with a random forest algorithm

Acknowledgements: Yolandi Le Roux; University of Pretoria

Predicting Delivery Times

- Location
- Driver
- Traffic
- Weather
- Residence type
- Order $
- Socio-demographics
- Distance

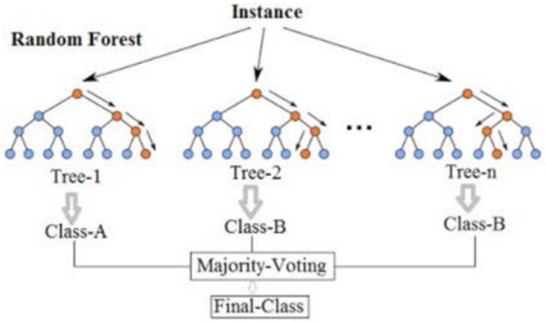


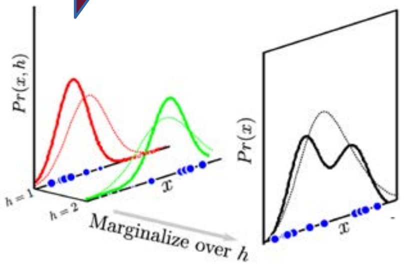


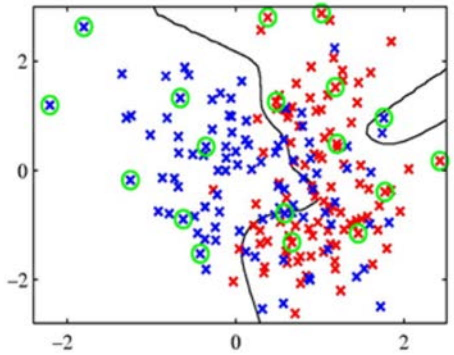
Customer Segmentation by means of Data Science

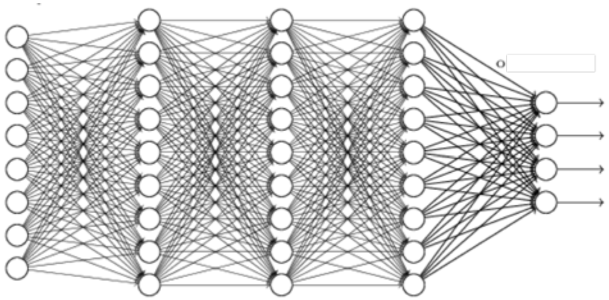
UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY
100
1918·2018

- Order qty
- Location
- Credit history
- $ contribution
- Payment type
- Products
- Complaints
- Socio-demographics
- Industry
- Cost 2 serve









Unique product service agreements

Business rules

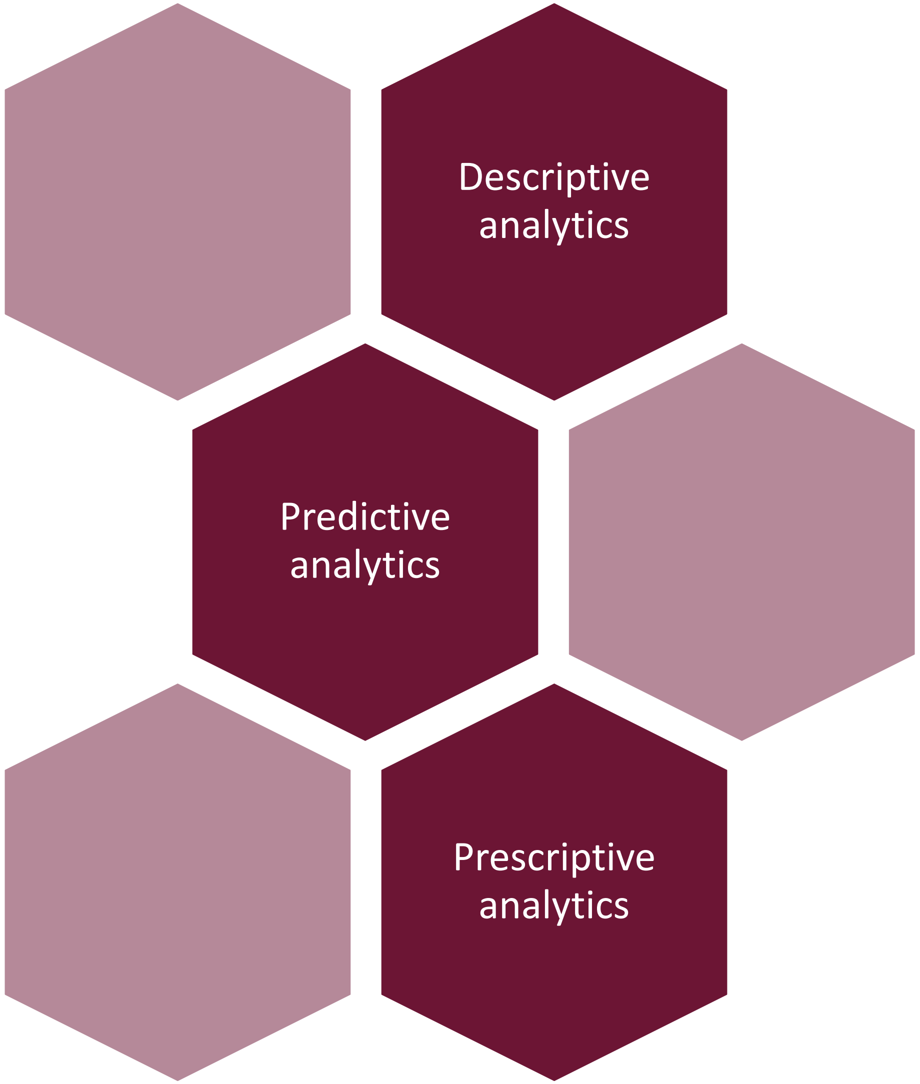
Resource planning

Other Projects

- Developing a recommender system for a wellness program
- Predicting port delays from wind, wave and other data
- Predicting manufacturing system quality
- Predicting energy requirements in the hospitality industry

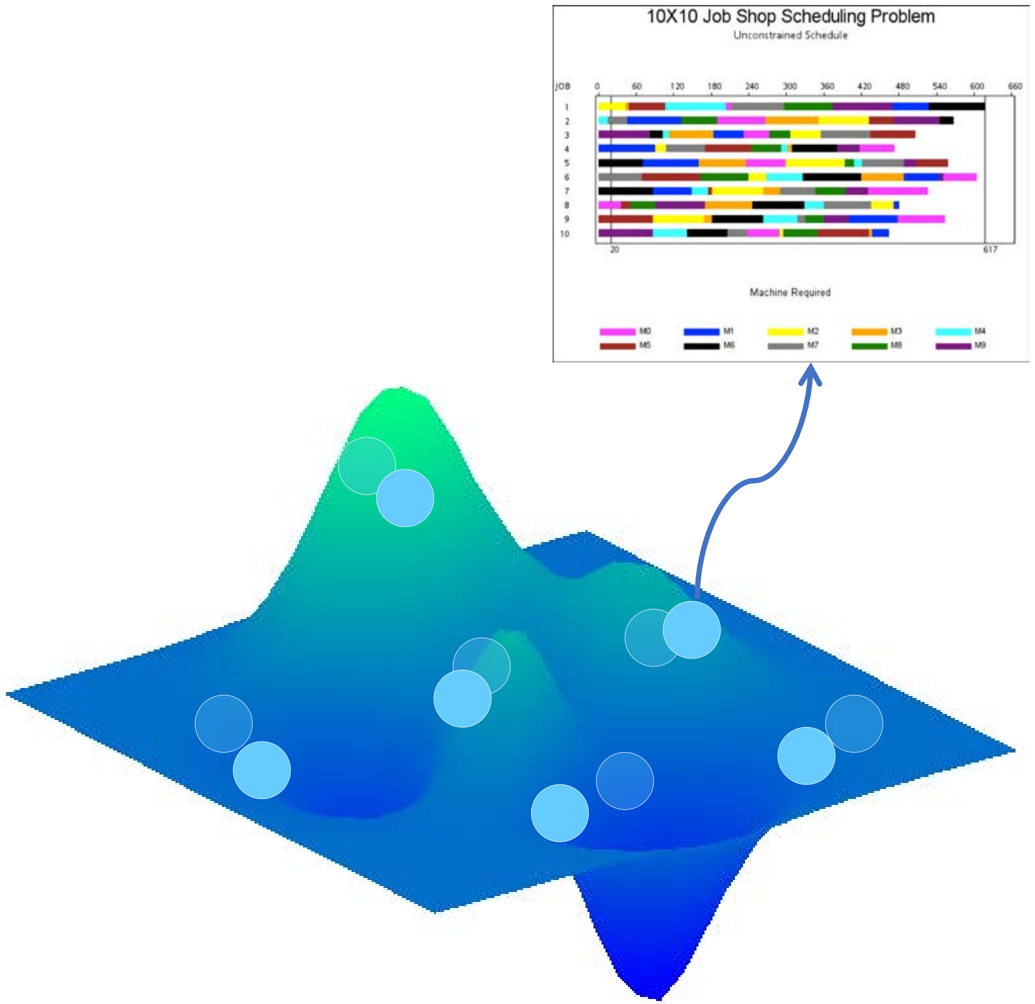
Acknowledgements: Lumi Dreyer; Andries Engelbrecht; Ridhaa Beneveld; Sibusiso Khoza; Cecil Musisinyani; Philip du Plooy; University of Pretoria

Data Science



Prescriptive Analytics

- Objective function
- Decision variables
- Constraints



Scheduling

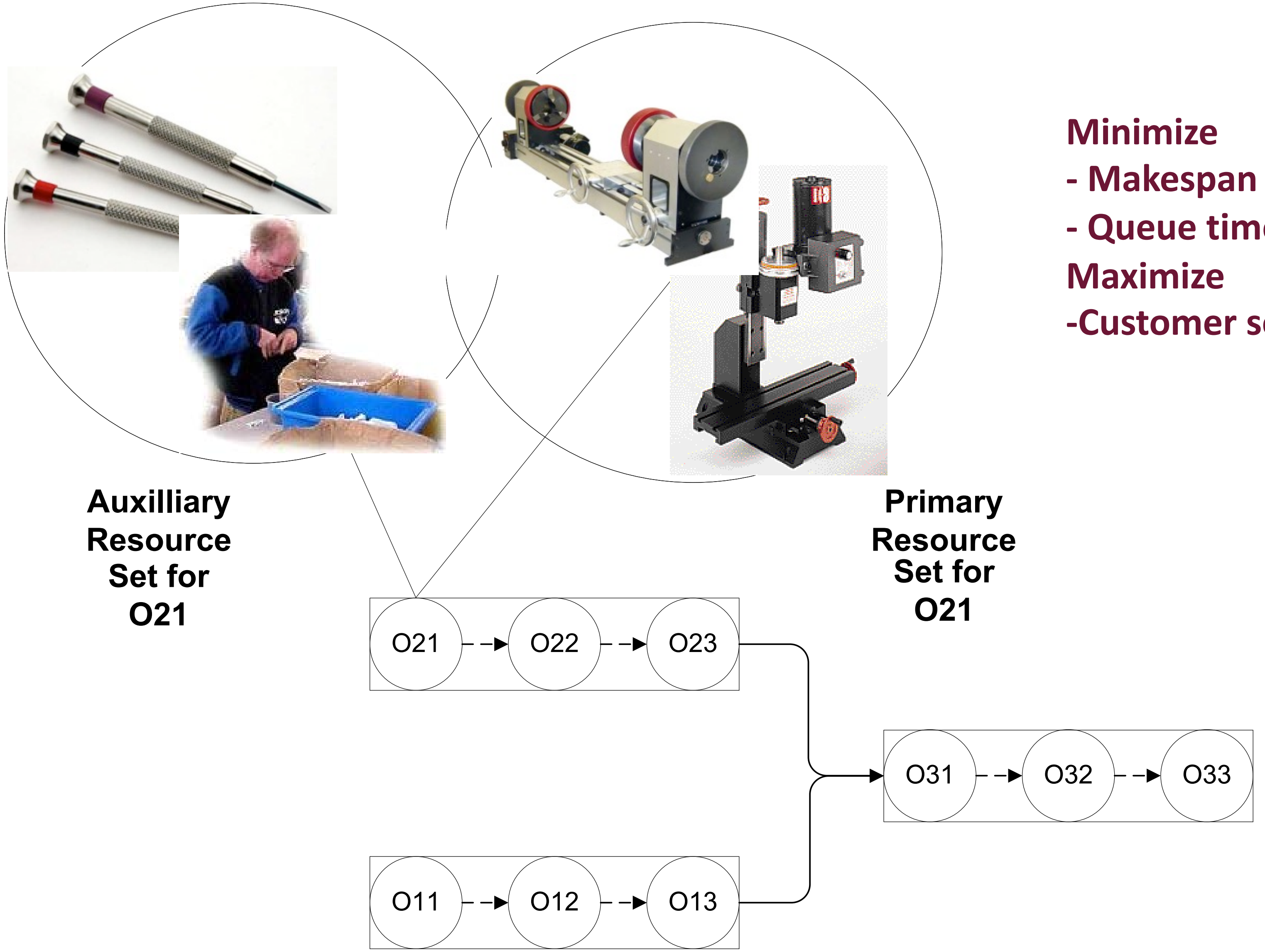
Operation 5

Operation 6

Operation 7

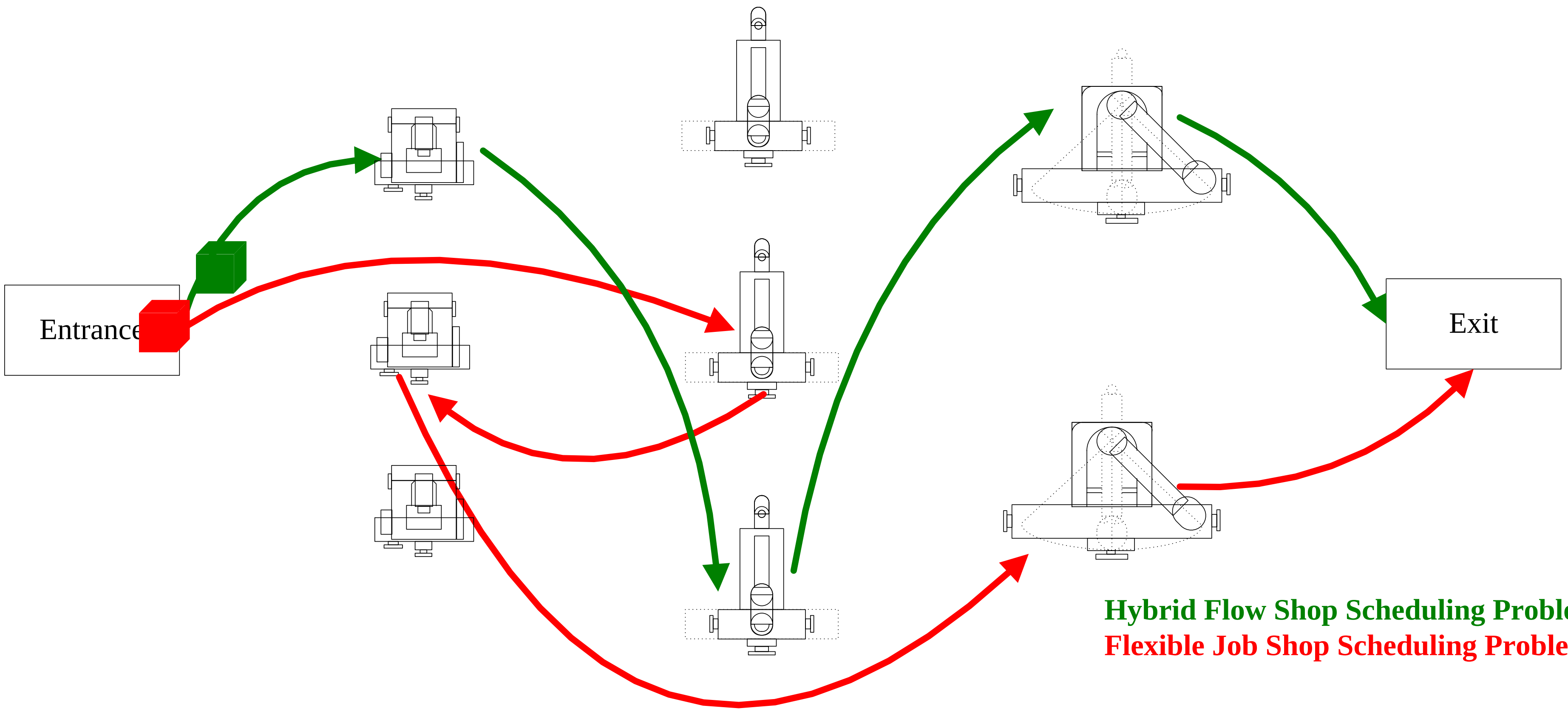
Resources	Time (Days)													
	12	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*
1	Operation 1				Operation 3			Operation 4						
2	Operation 2													
3								Operation 4						
4	Operation 1													

Scheduling



Minimize
- Makespan
- Queue time
Maximize
-Customer service

“Multiple Resources per Operation” Models



Hybrid Flow Shop Scheduling Problem
Flexible Job Shop Scheduling Problem

Optimization Literature Example

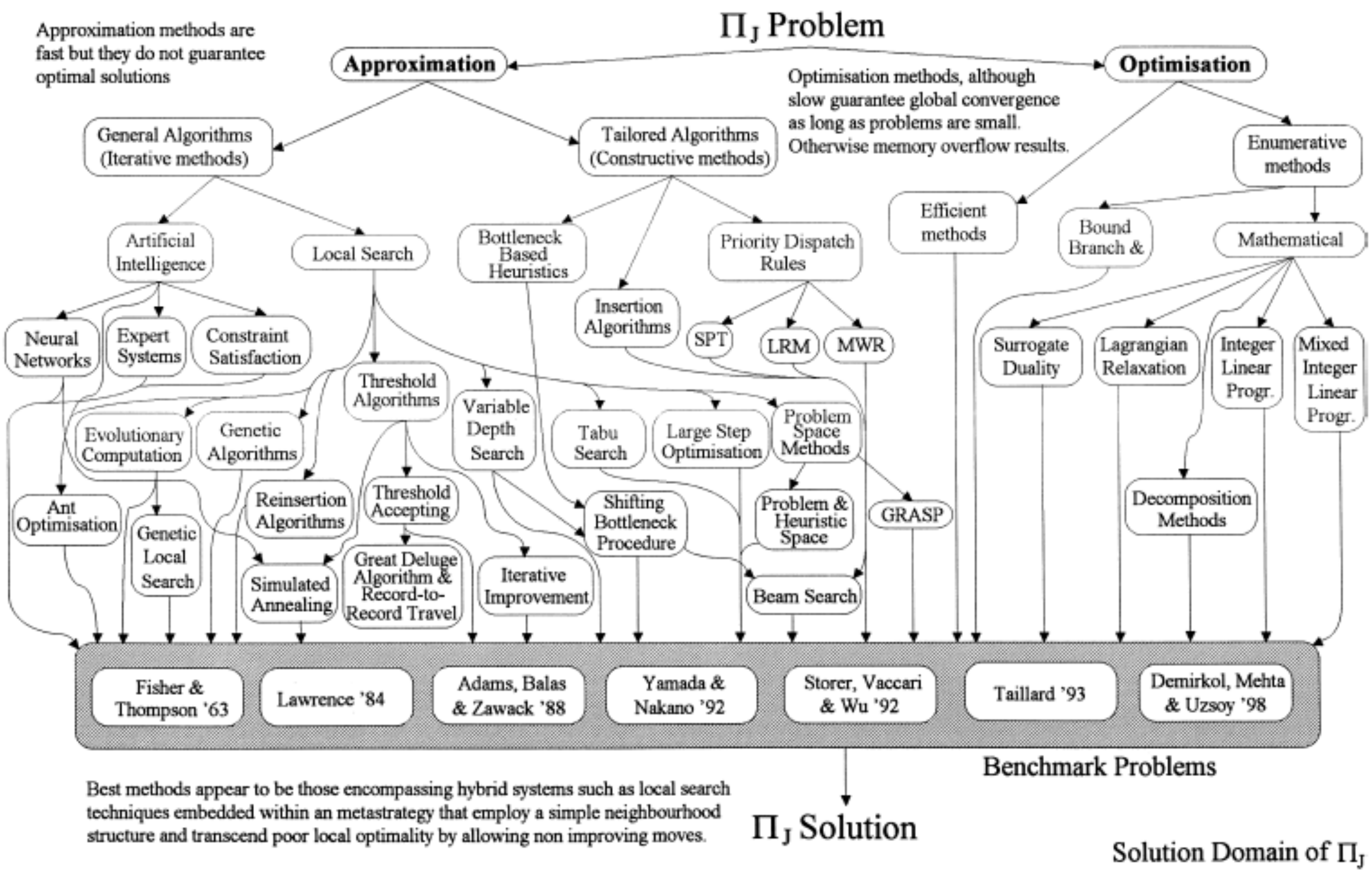
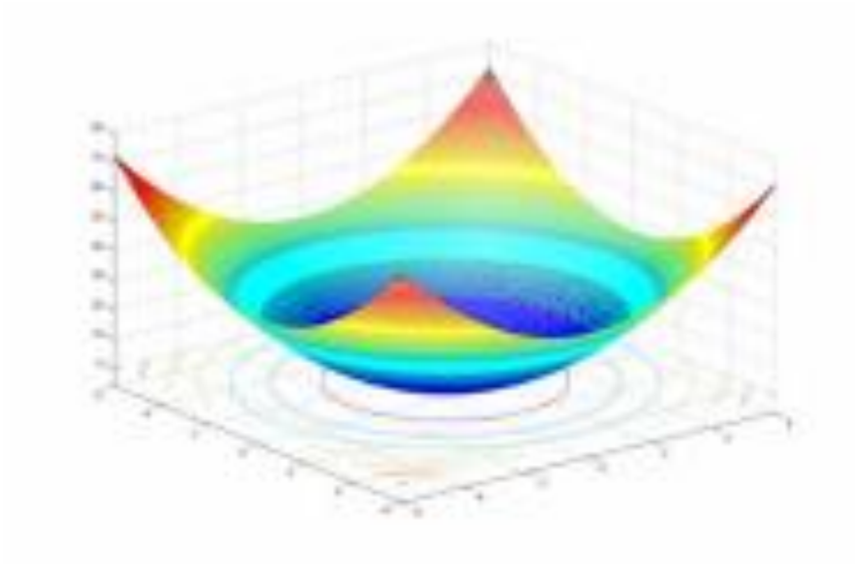


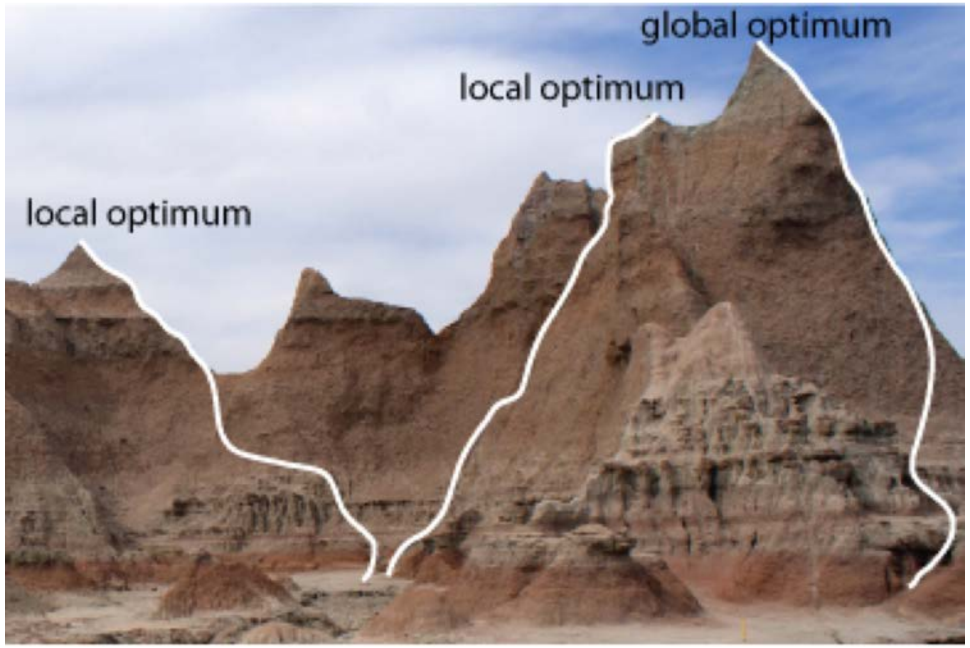
Fig. 1. The phases of Π_J research.

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY
100
1918 · 2018

The Local Search Algorithm



f(x)



x

Source: http://community.asdlib.org/imageandvideoexchangeforum/

But why do we have to pick one algorithm?

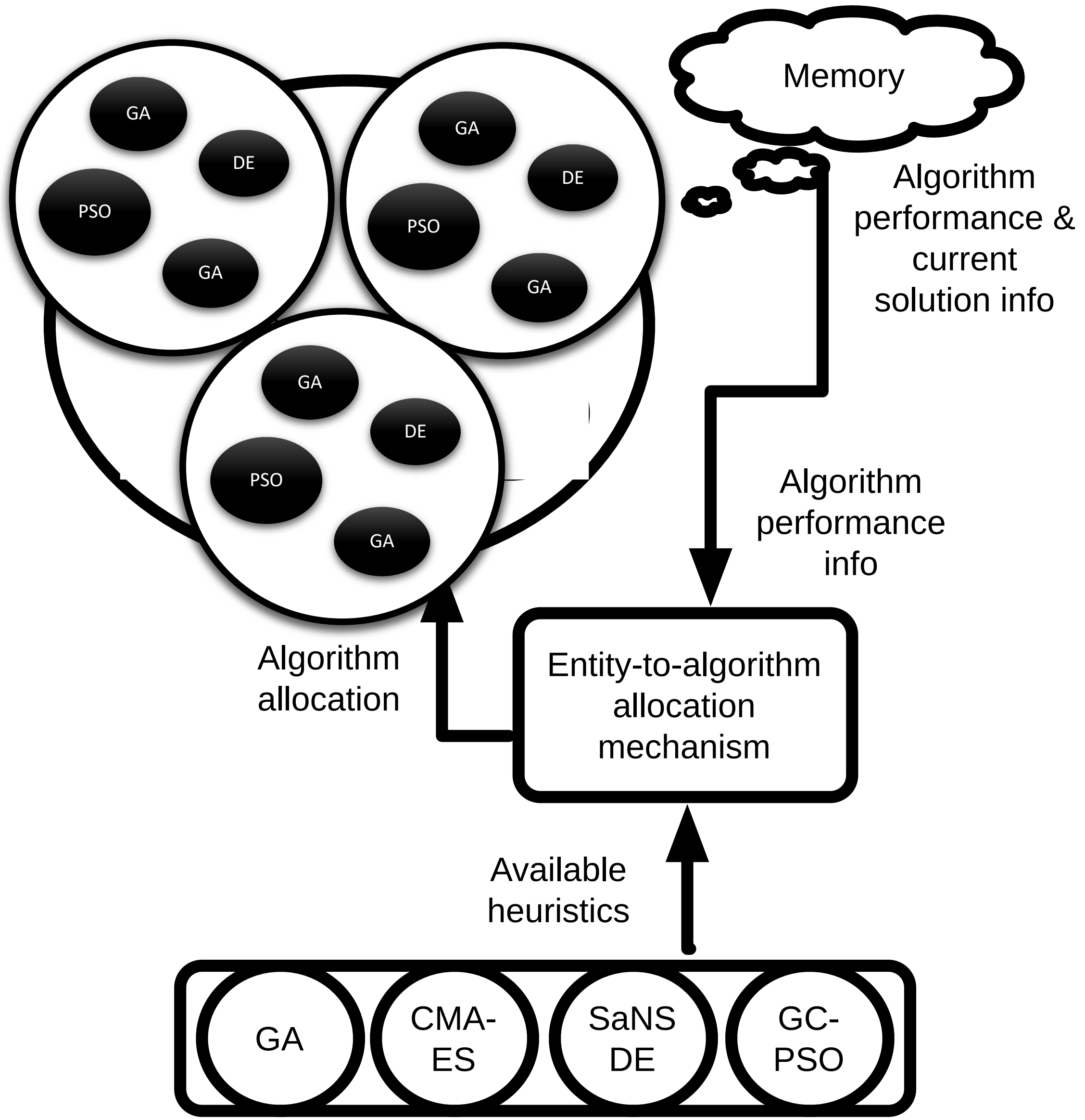
and

Why can’t we let the algorithm pick the algorithm?

…

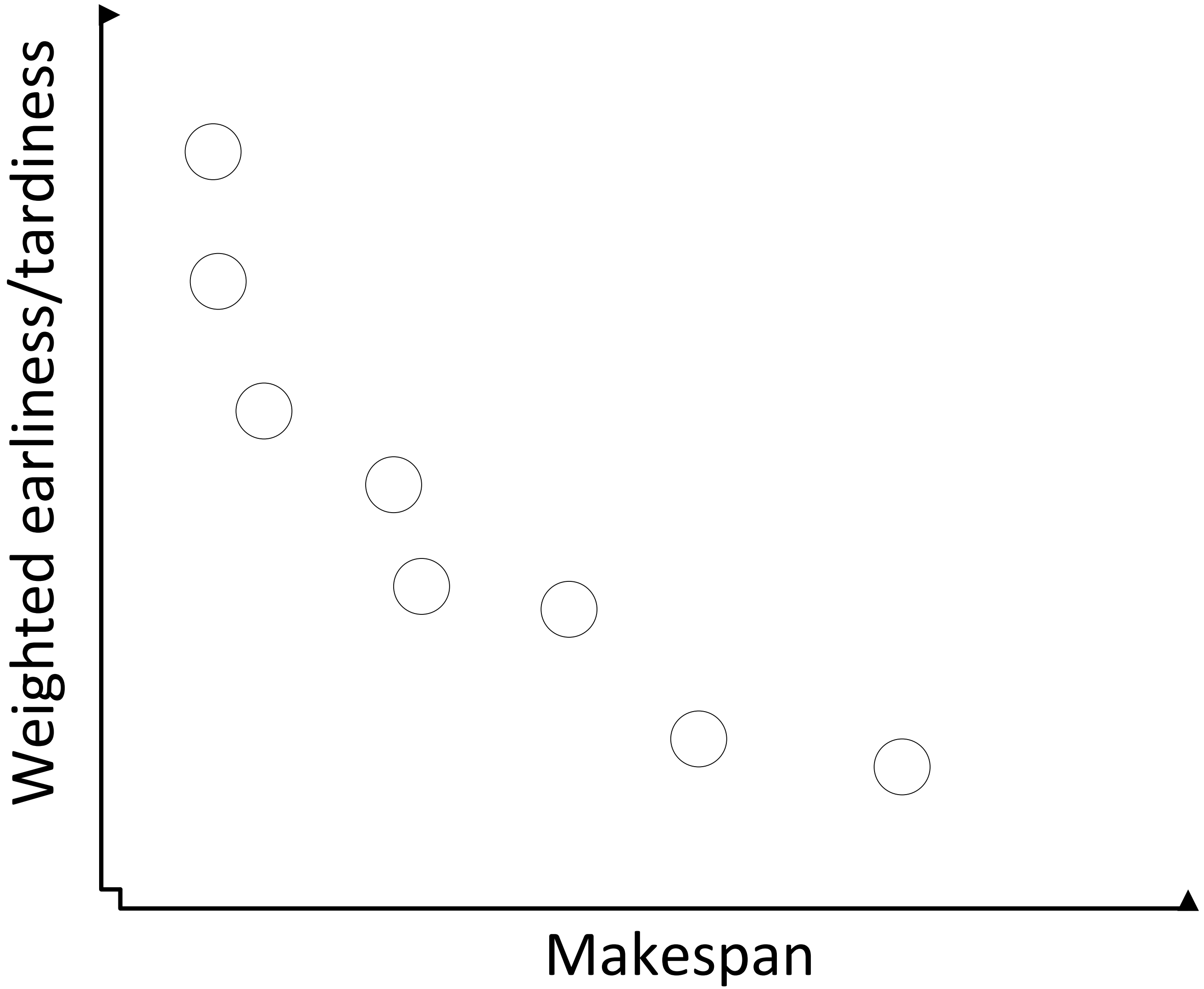
- 24 algorithm variations
- 60 diverse problems
- 24 000 hrs of processing time

The HMHH scheduling algorithm

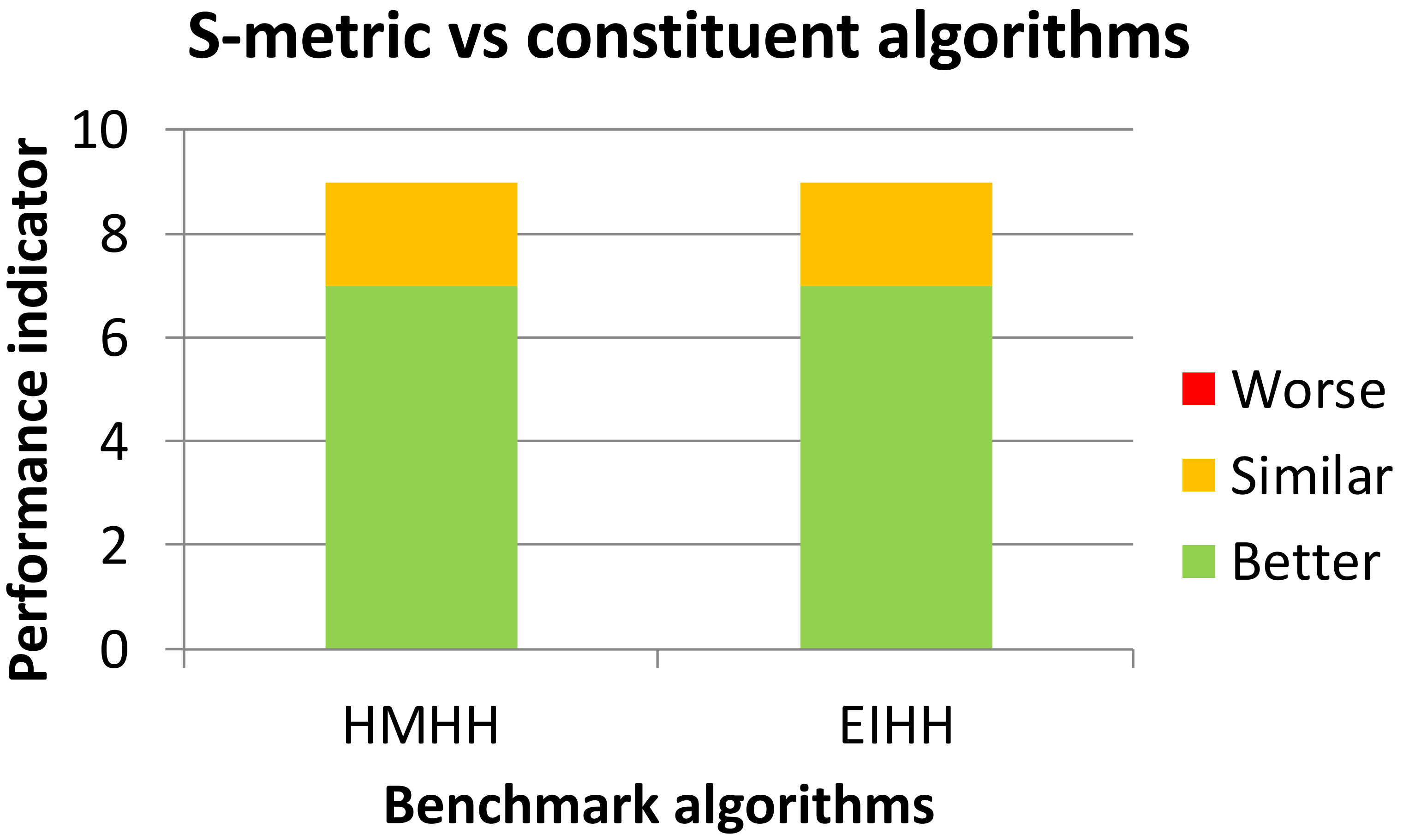


UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY
100
1918·2018

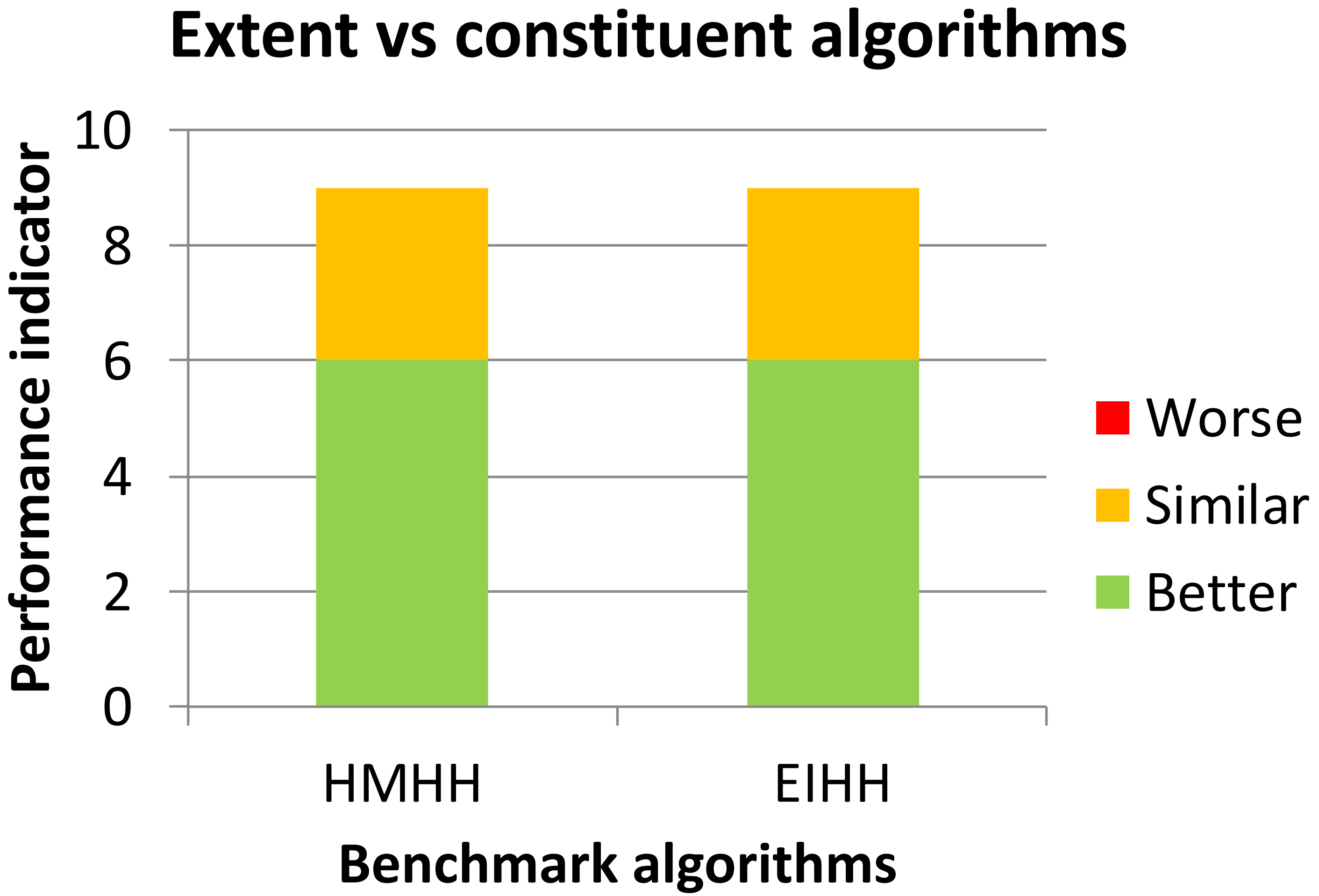
Multi-objective scheduling



Results interpretation



Results interpretation



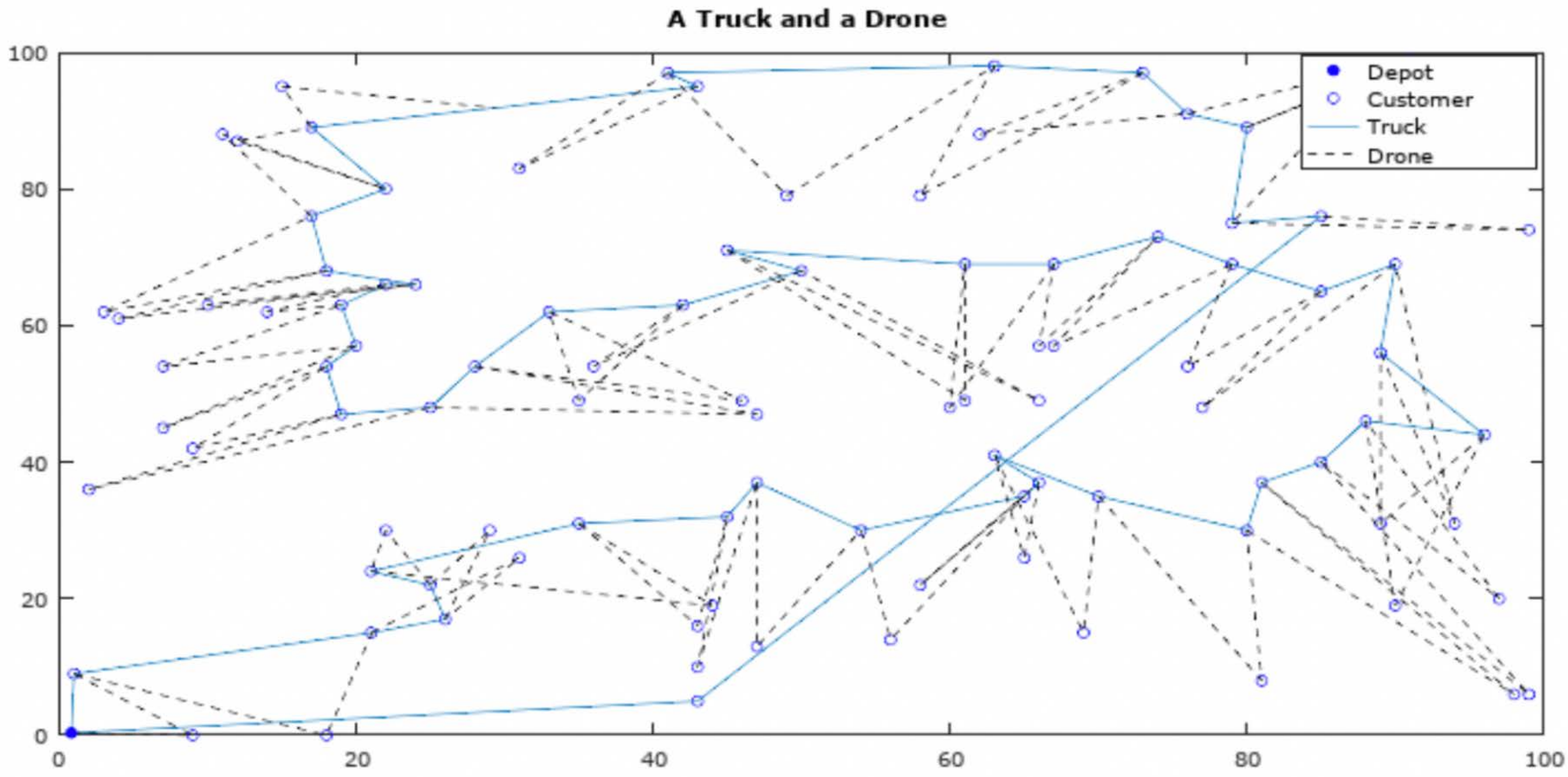
Impact

- 68% improvement when compared with state-of-the-art algorithms

- At 40 hrs per design experiment and eliminating 5 experimental designs, you save 220 development hours per algorithm application


amazon

Preliminary Results



Acknowledgements: Tsietsi Moremi


DATA

Data science:
How to leverage data for superior decision making

Jacomine Grobler
jacominegrobler@sun.ac.za