

Automatic Multi-Objective Modeling with Machine Learning

Prasanna Balaprakash

Mathematics and Computer Science Division &
Leadership Computing Facility
Argonne National Laboratory

Argonne Training Program on Extreme-Scale Computing (ATPESC)
Aug 5th, 2017

Outline

Motivation

AutoMOMML

Applications

Summary & Future

Application landscape

Argonne's mission: science-based solutions to pressing global challenges through discovery and transformational science and engineering...

World-leading
hard x-ray
sciences and
sources

Discovery
science for
energy

Leadership
computing and
computational
ecosystem

Fundamental
physics and
accelerator
capabilities

Materials &
systems
engineering
solutions

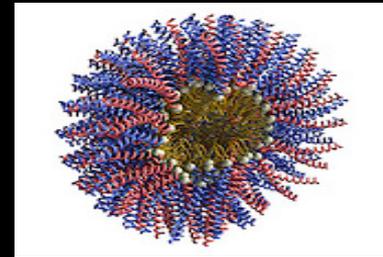
Energy Storage

Sustainable
Transportation

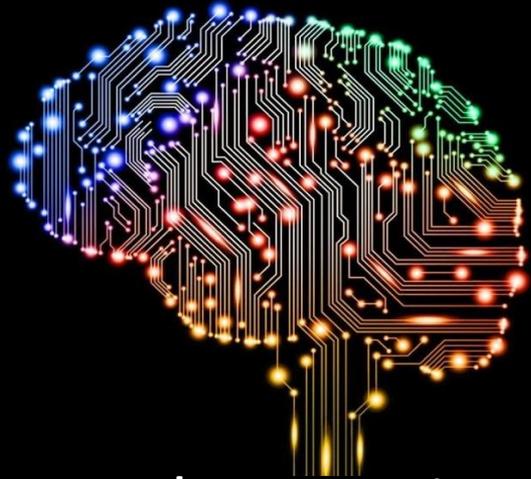
Nuclear Energy

Environmental
Genomics

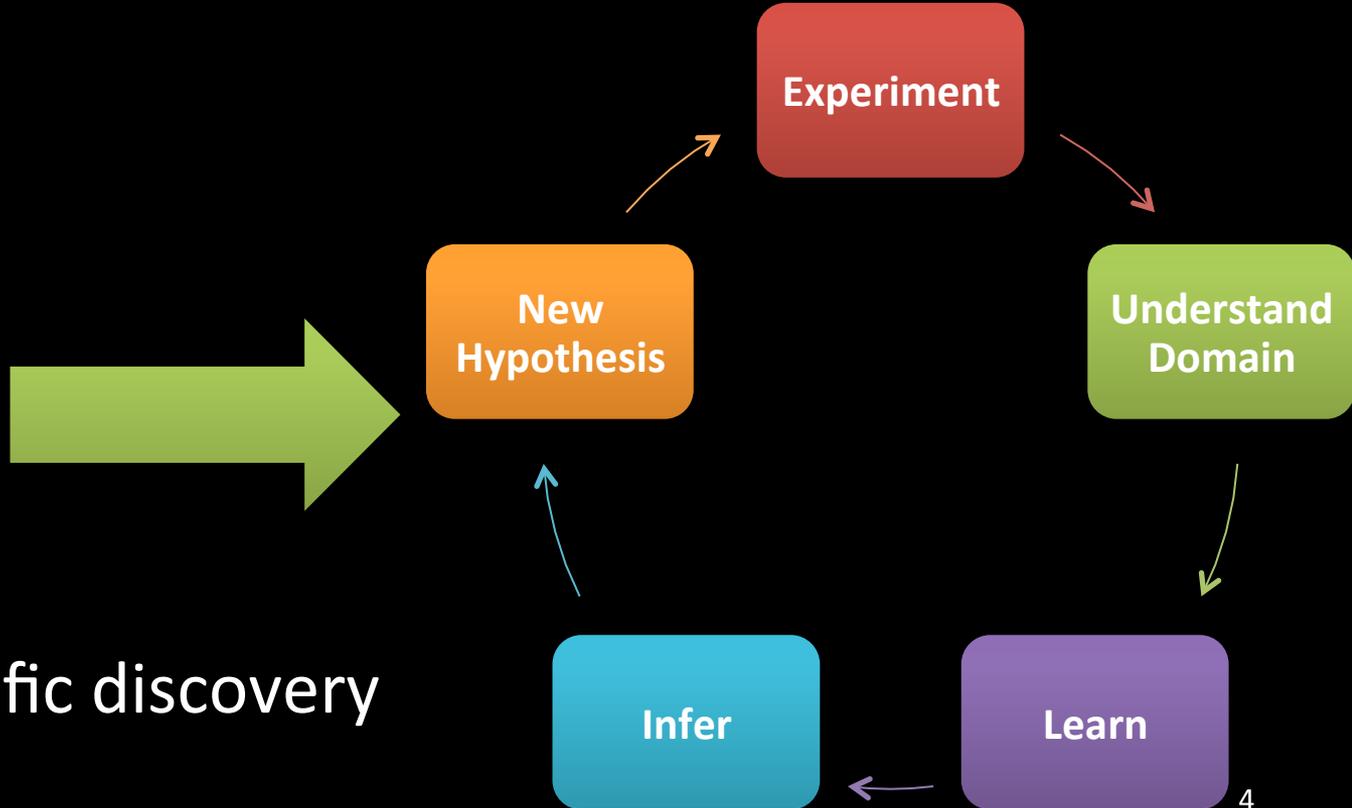
National
Security

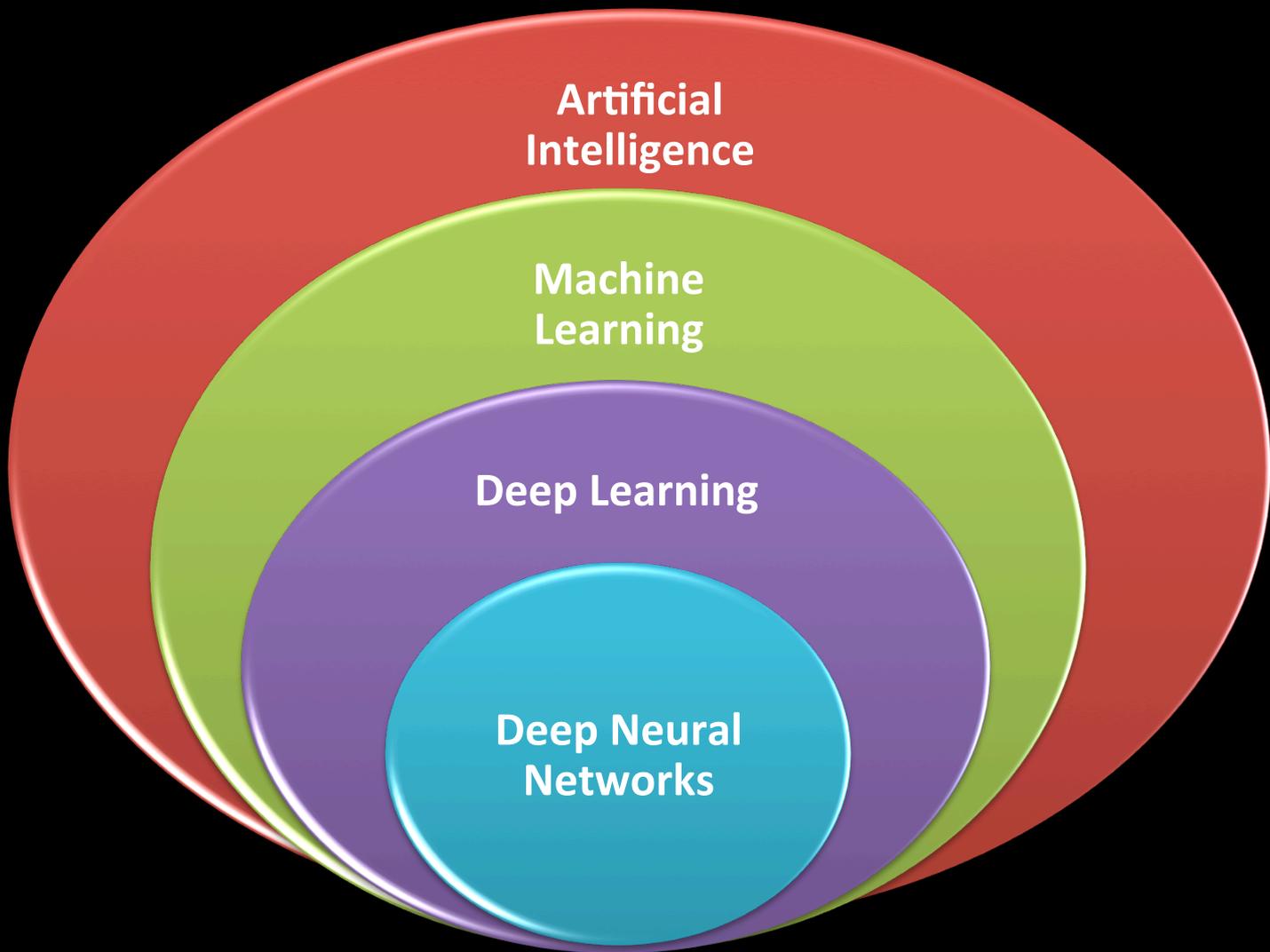


Artificial Intelligence for science



Accelerate scientific discovery





**Artificial
Intelligence**

**Machine
Learning**

Deep Learning

**Deep Neural
Networks**

Machine Learning

Unsupervised learning



Reinforcement learning



Supervised learning



Optimization and statistics



Supervised learning

Inputs **Outputs**



Cat



Dog



Horse

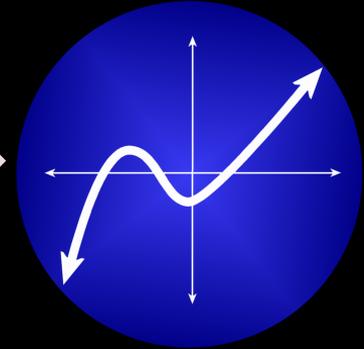
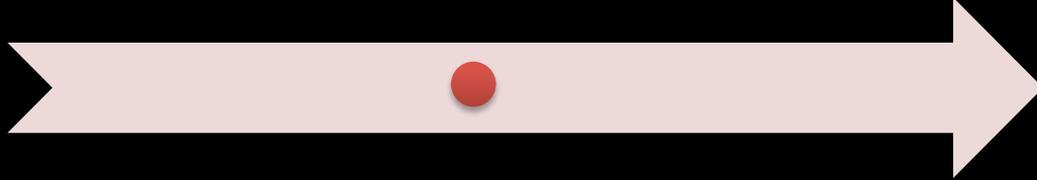


Elephant

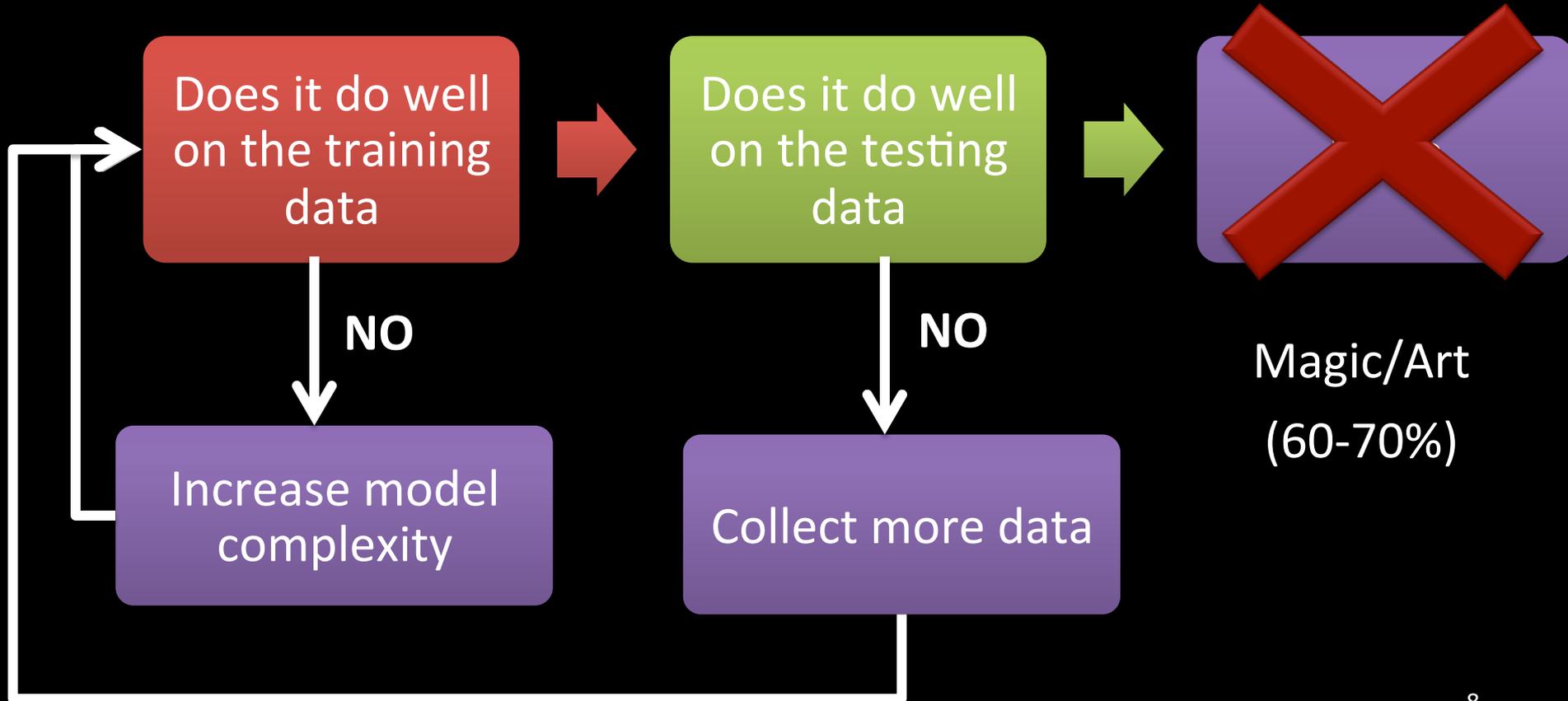


Tiger

Training



Supervised learning





Nvidia To Train 100,000 Developers In 'Deep Learning' AI To Bolster Healthcare Research



PODCAST: The Limit Does Not Exist: How To Be Everything

MAY 11, 2017 @ 01:49 PM 2,765 

The Little Black Book of Billion

Nvidia To Train 100,000 Developers In 'Deep Learning' AI To Bolster Healthcare Research



Lee Bell, CONTRIBUTOR

[FULL BIO](#) 

Opinions expressed by Forbes Contributors are their own.

Artificial Intelligence (AI) pioneer Nvidia has announced it will train 100,000 developers in "deep learning" to bolster health care research and improve treatment in diseases like cancer.

Intelligent Machines

Why Google's CEO Is Excited About Automating Artificial Intelligence

AI software that can help make AI software could accelerate progress on making computers smarter.

by Tom Simonite May 17, 2017

Outline

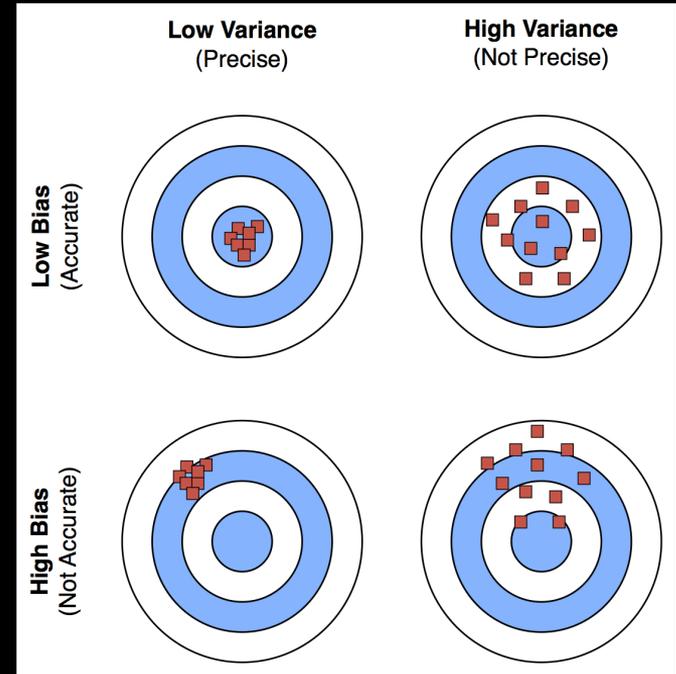
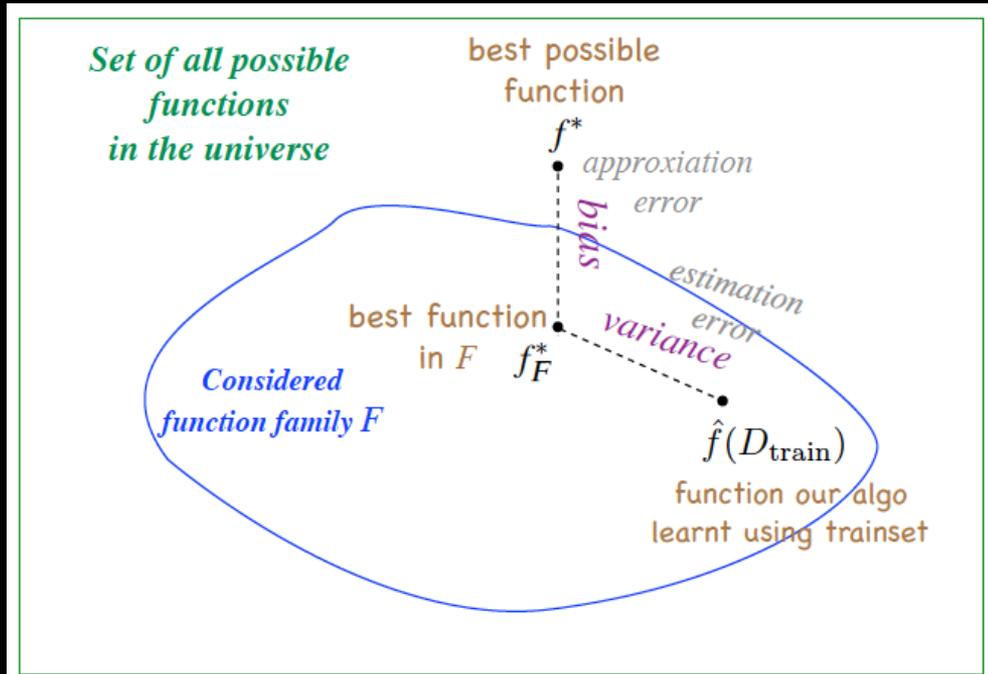
Motivation

AutoMOMML

Applications

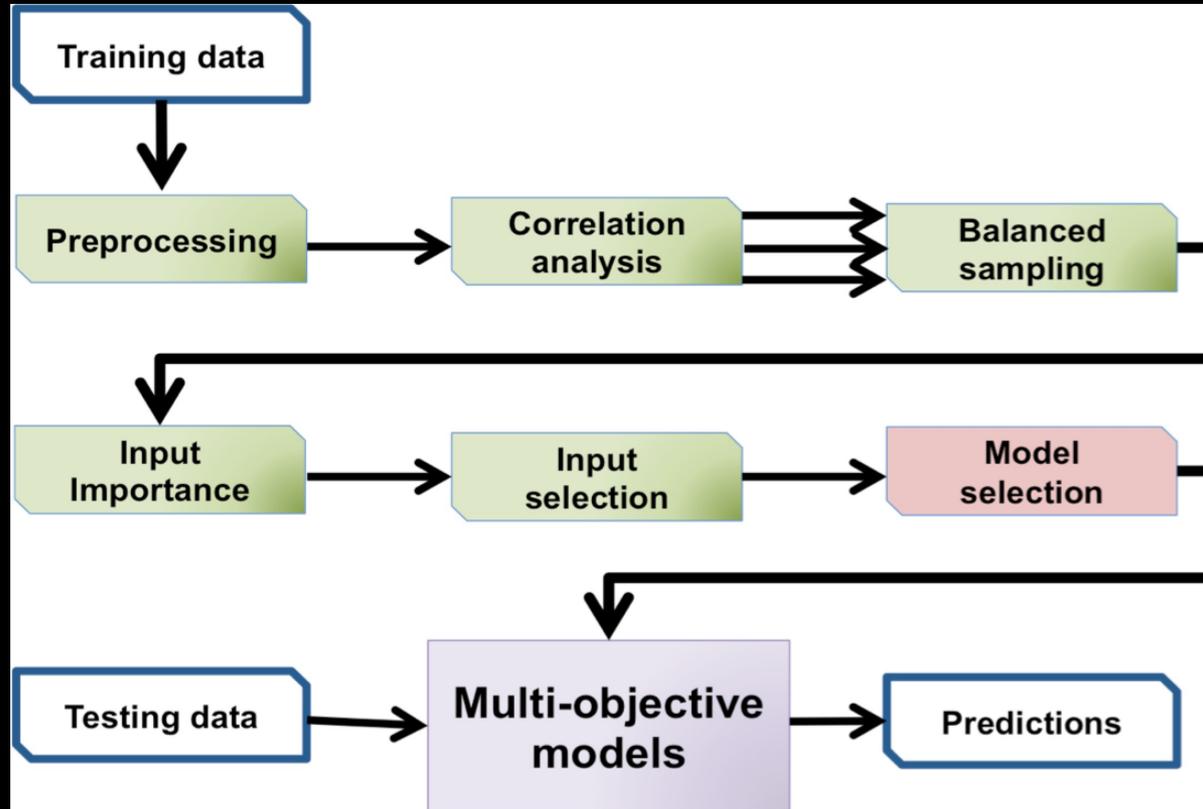
Summary & Future

Bias variance tradeoff

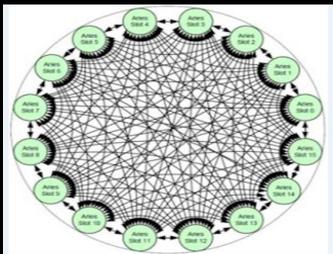
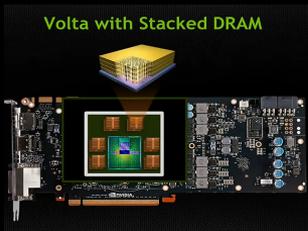
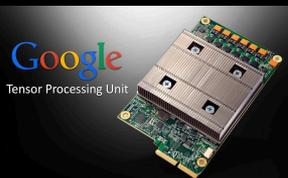


- Learning algorithms *seek to reduce bias and variance* in a different way
- **No free lunch**: no single algorithm will work well on all data set

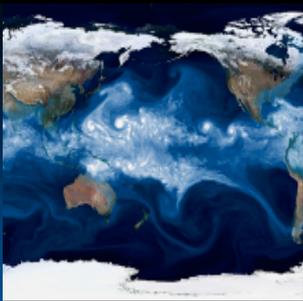
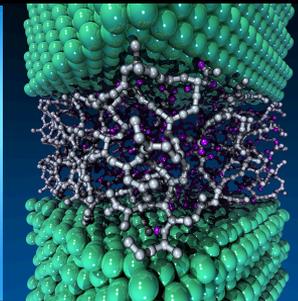
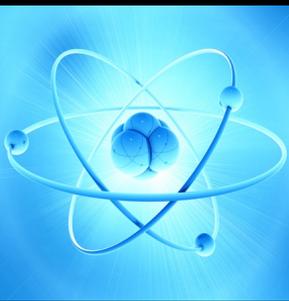
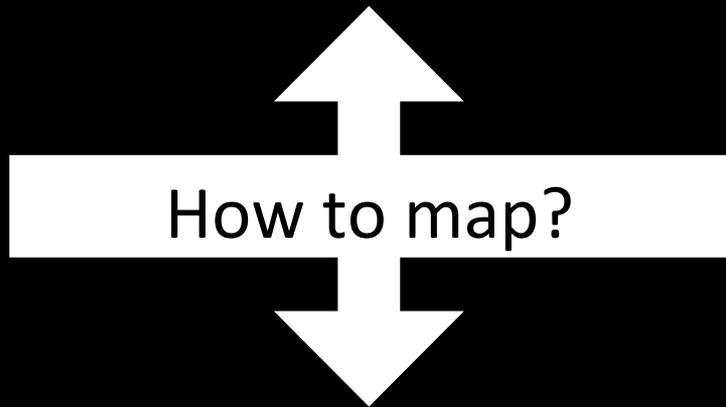
AutoMOMML: Automatic Multi-Objective Modeling with Machine Learning



Mapping problem in HPC

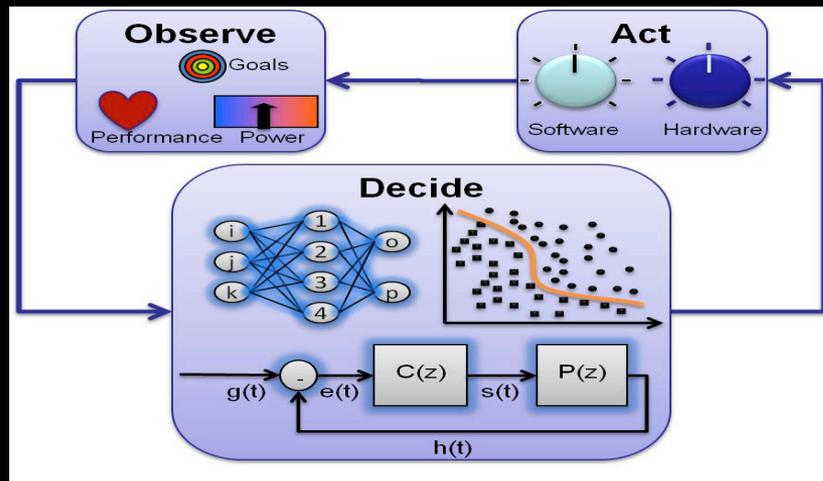
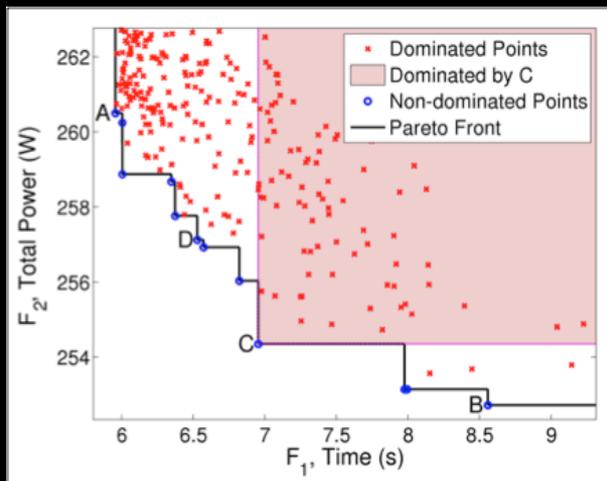


Diverse architectural landscape



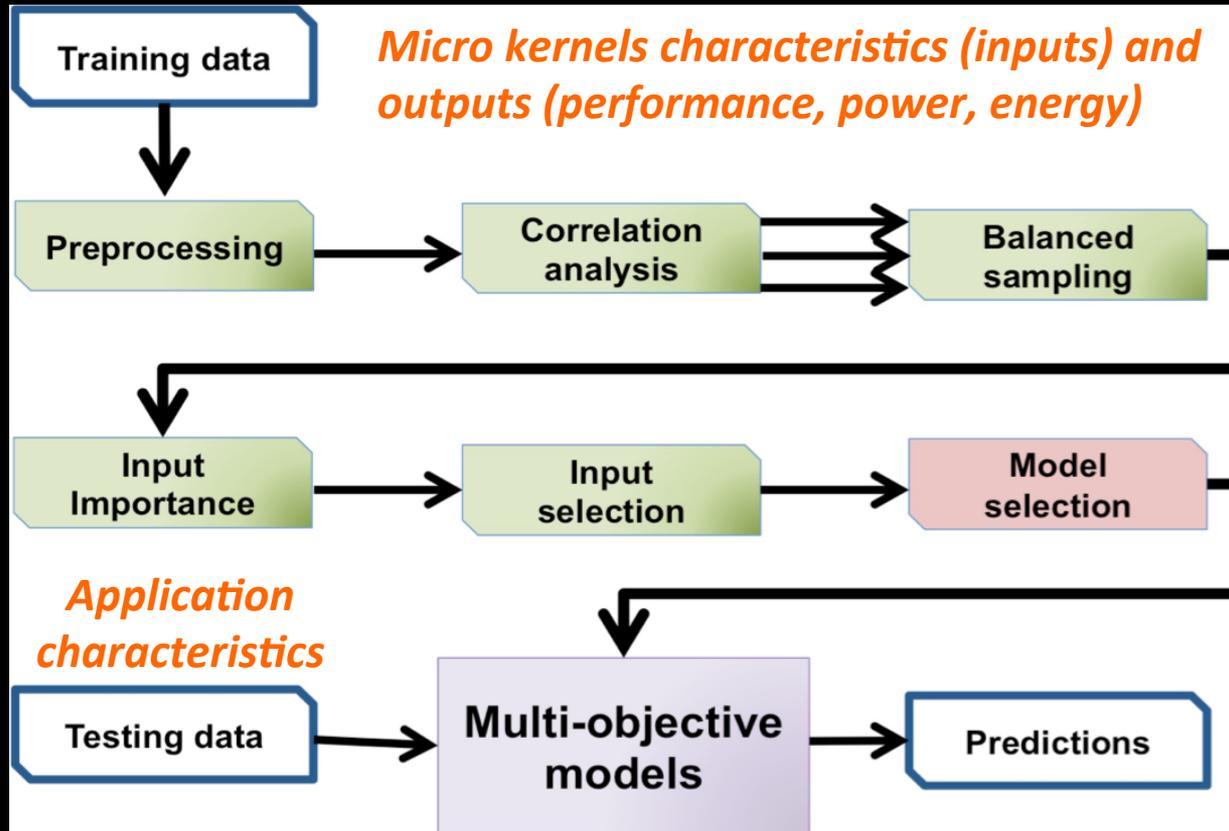
Diverse application landscape

Performance modeling in HPC

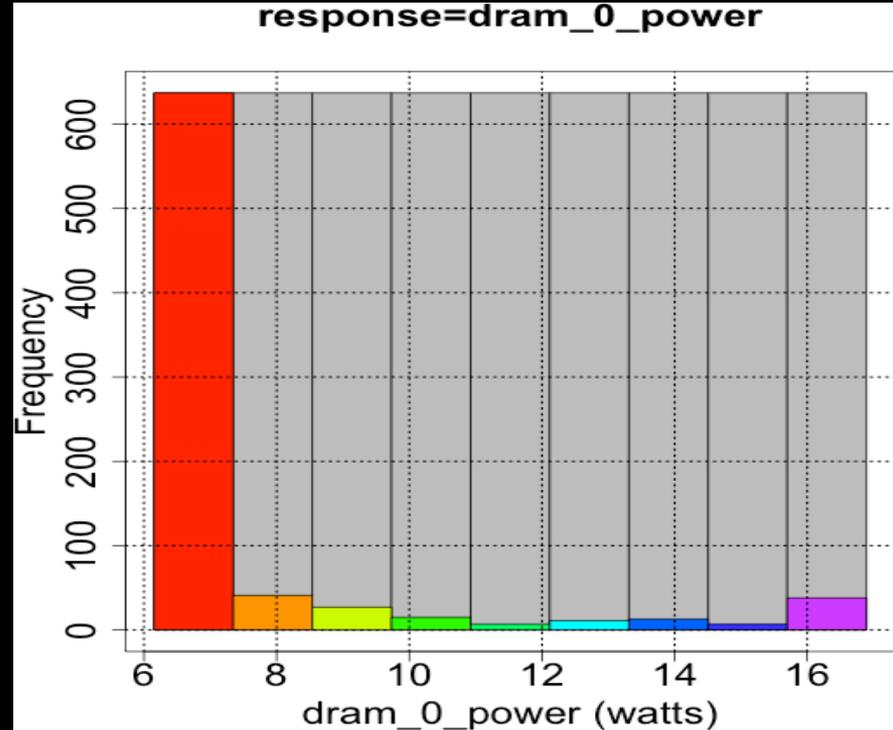
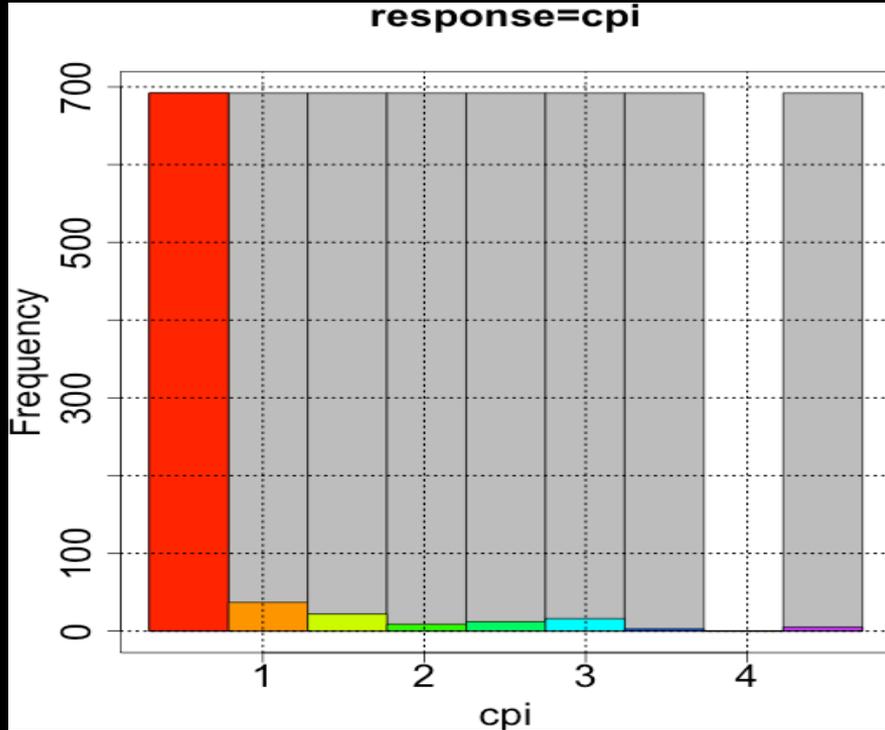


- Insights on important knobs that impacts performance
- Avoid running code configuration on the target machine
- Help prune large search spaces in performance tuning

AutoMOMML for HPC

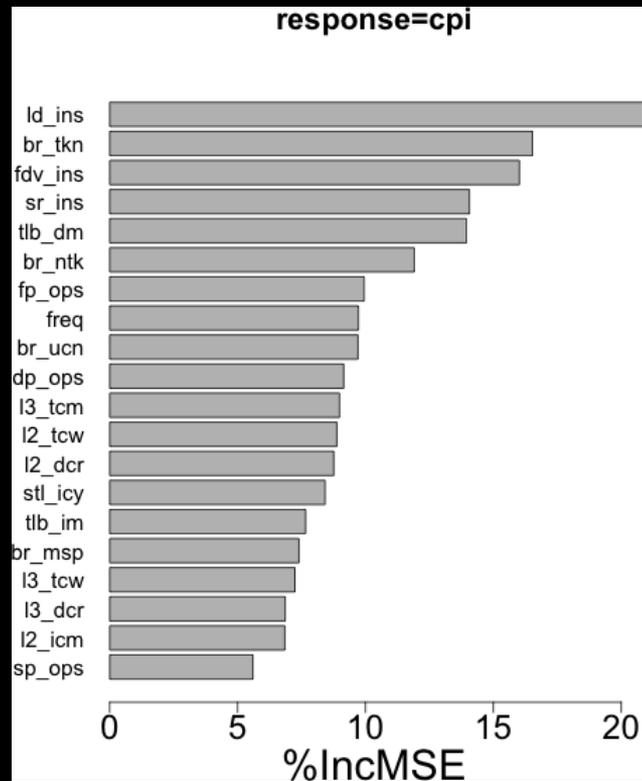
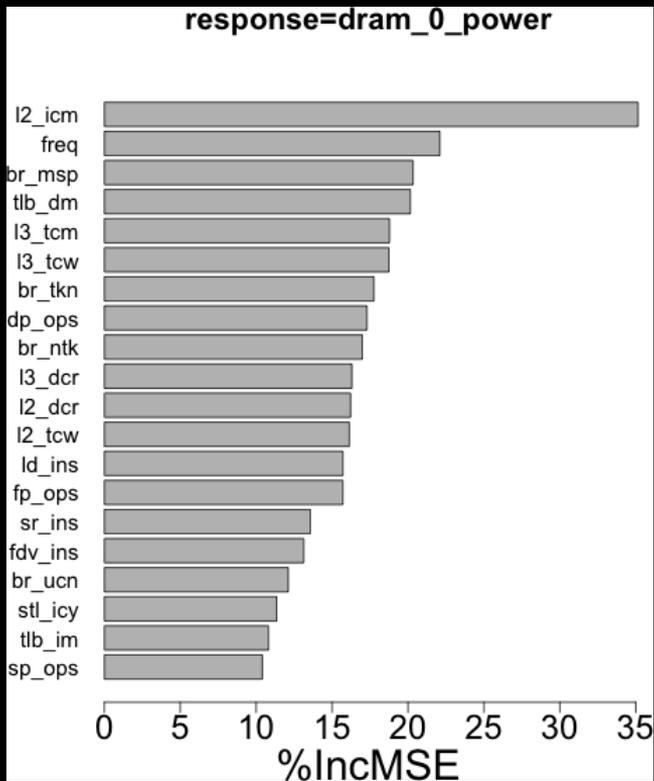


Balanced sampling



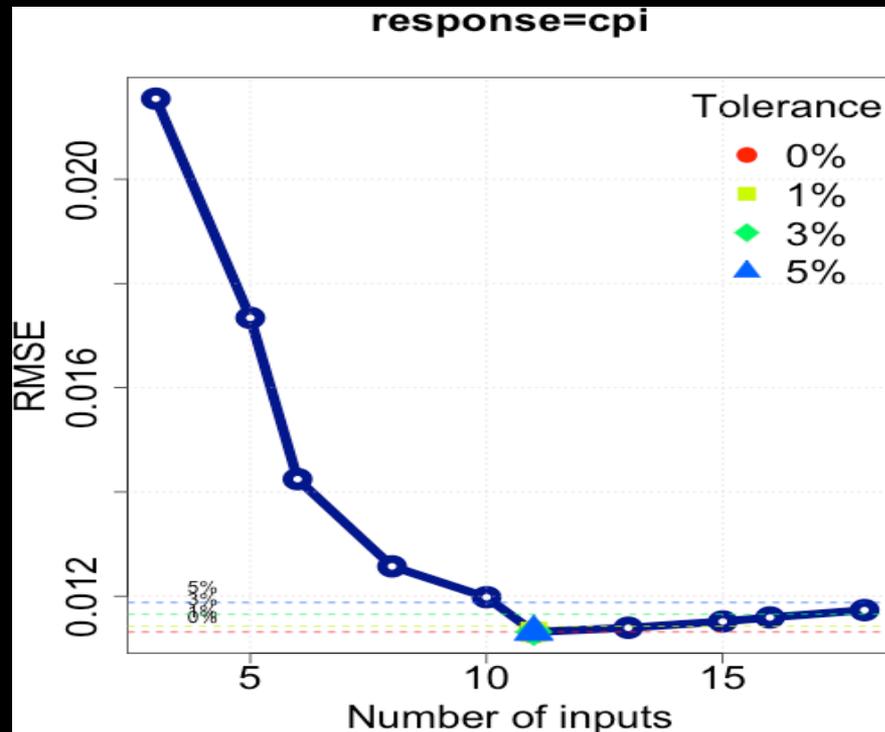
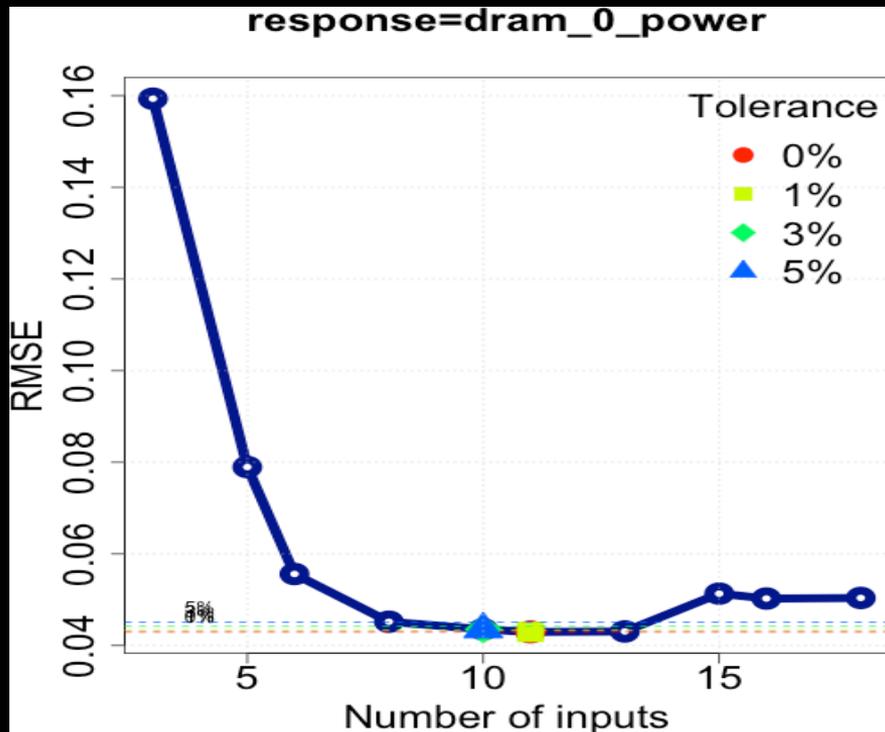
- Over-sampling to avoid bias in the training points
- Training points are sampled repeatedly from low-frequency ranges

Input importance via random forest



- Random forest's *permutation accuracy* for input importance
 - Randomly permute the values of a parameter to check the impact

Model-based input selection

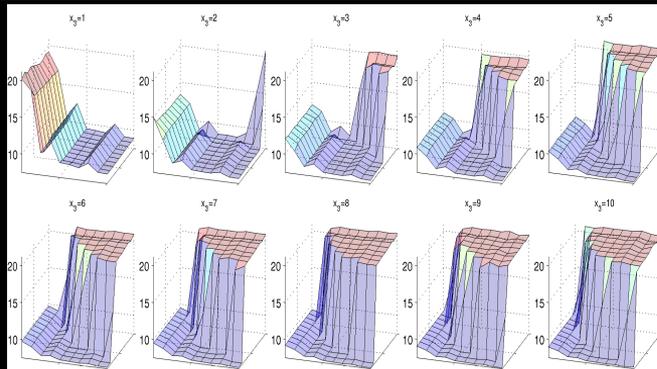


- Recursive feature elimination
- RMSE reduction becomes insignificant after a certain number of inputs

Hyperparameter tuning

Unevaluated parameter configurations

Optimization



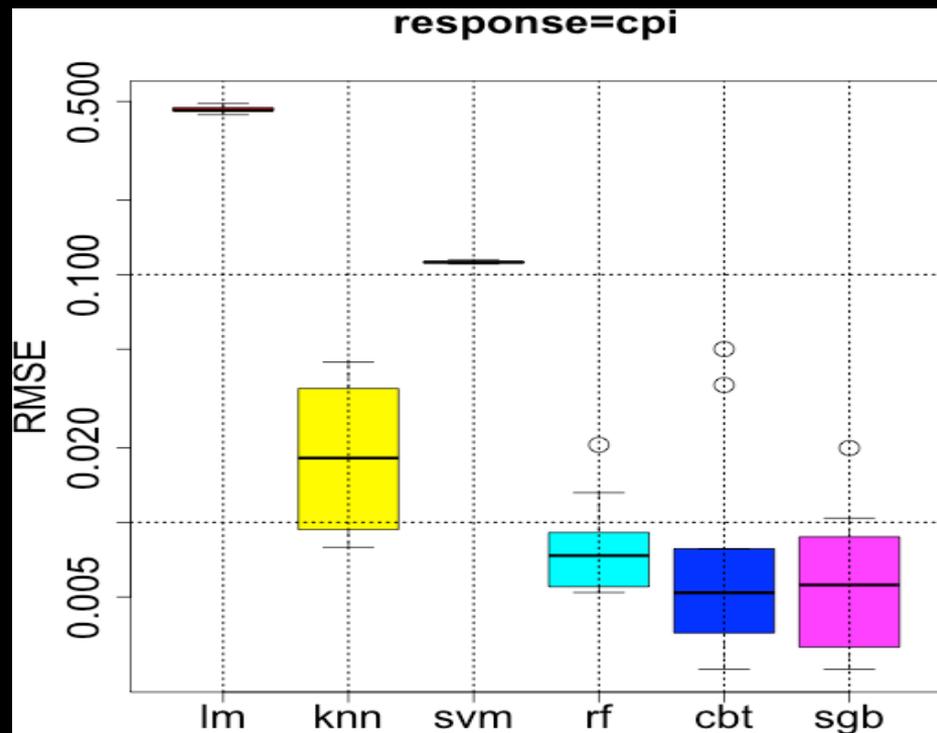
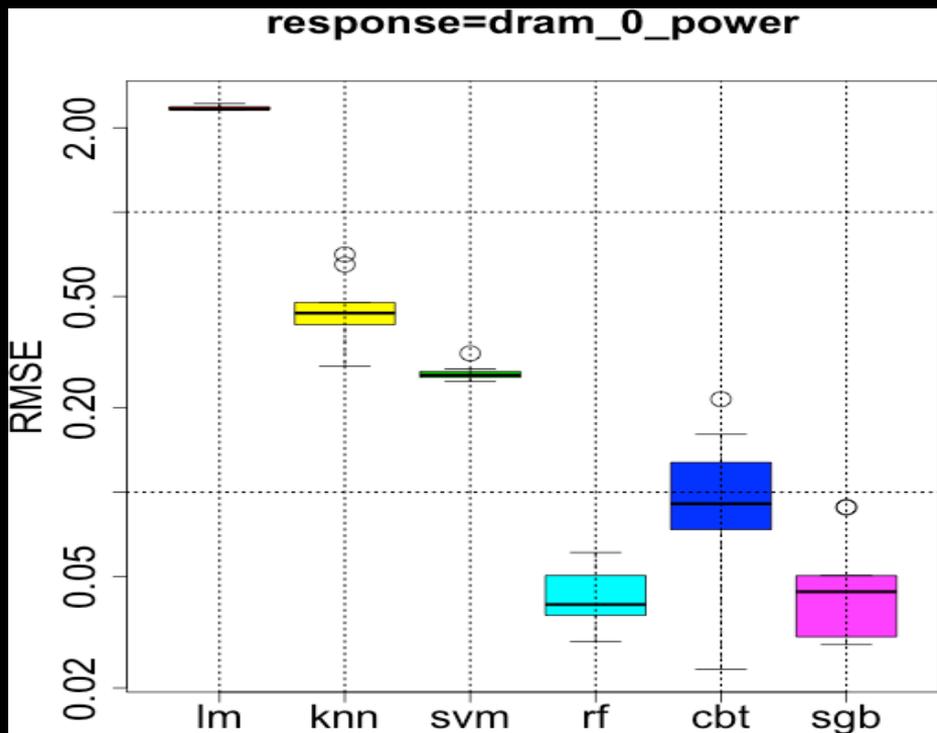
Performance metrics

Promising configurations

Evaluation

iterative refinement improves model performance

Model selection



- **t-test** establishes different model combinations based on given output

Outline

Motivation

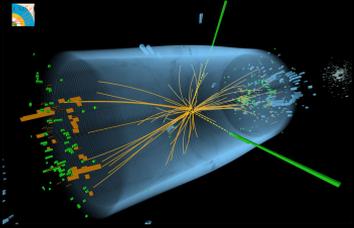
AutoMOMML

Applications

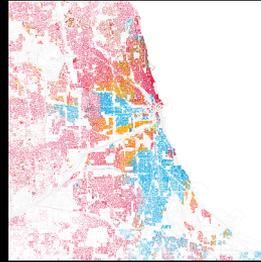
Summary & Future

AutoMOMML

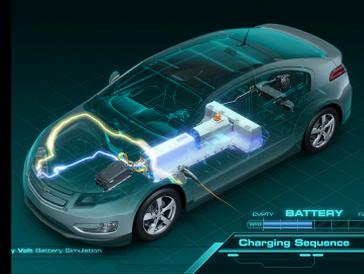
LHC & APS



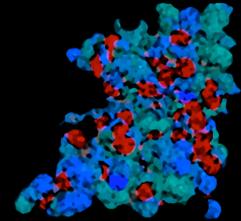
Urban Data Analytics



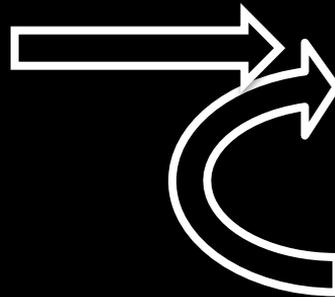
Vehicle Tech Assessment



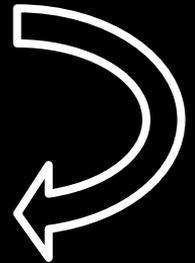
Drug Discovery



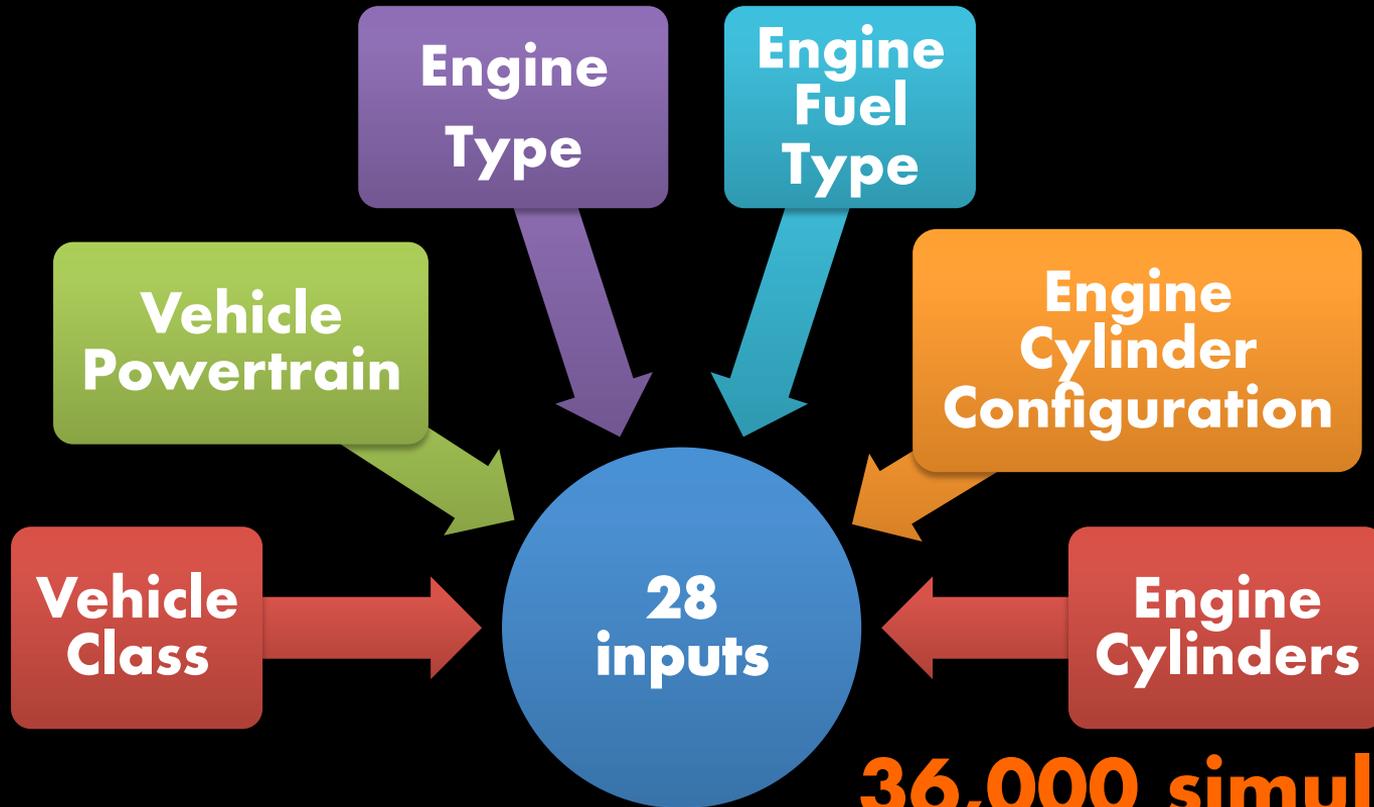
Search space of algorithms



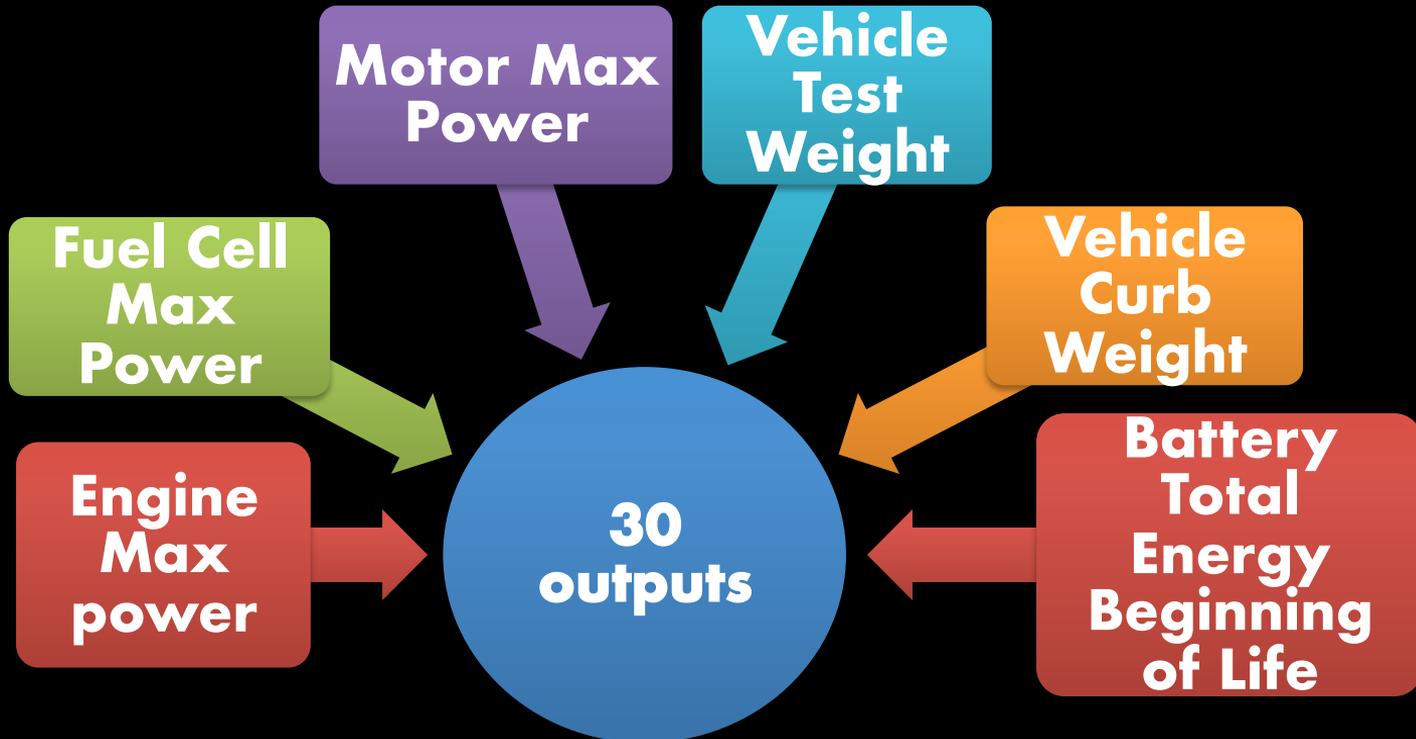
Intelligent navigation



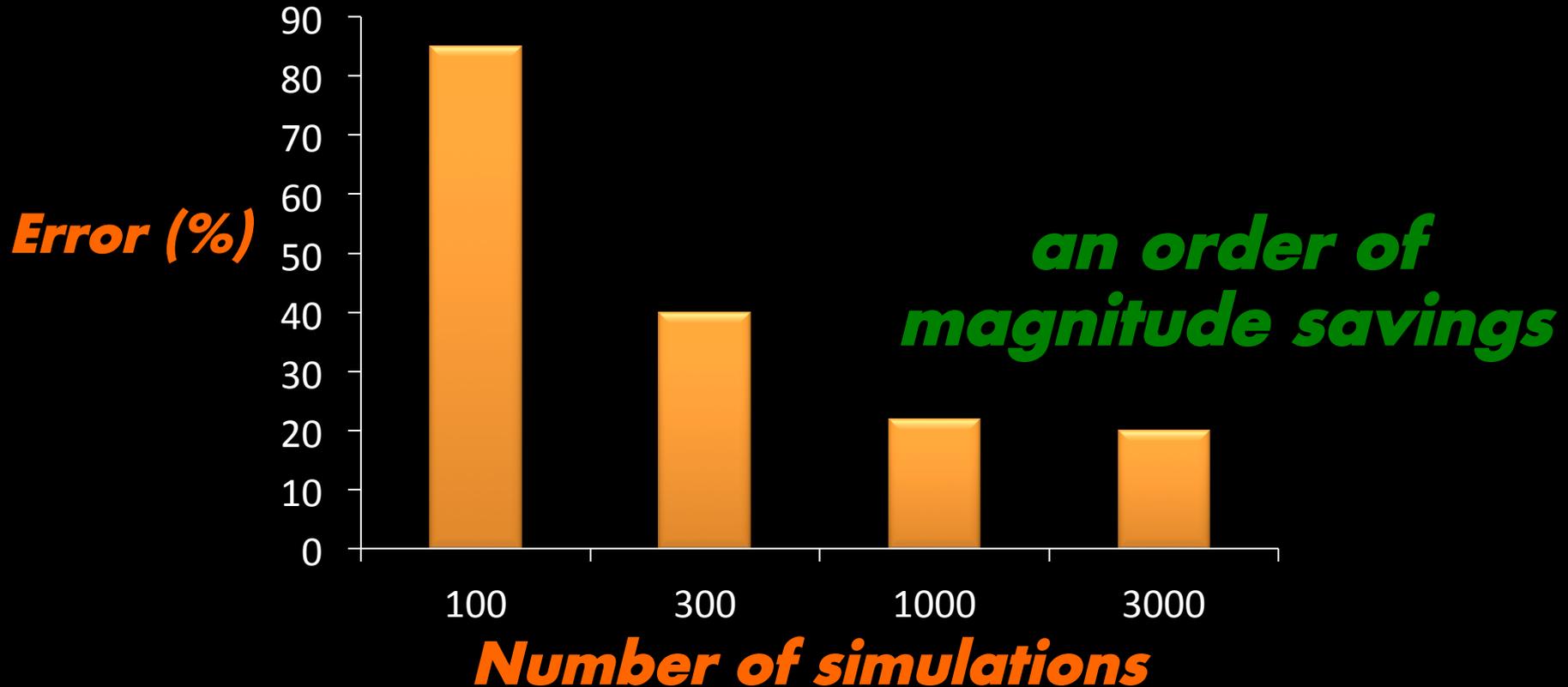
Simulation campaign reduction



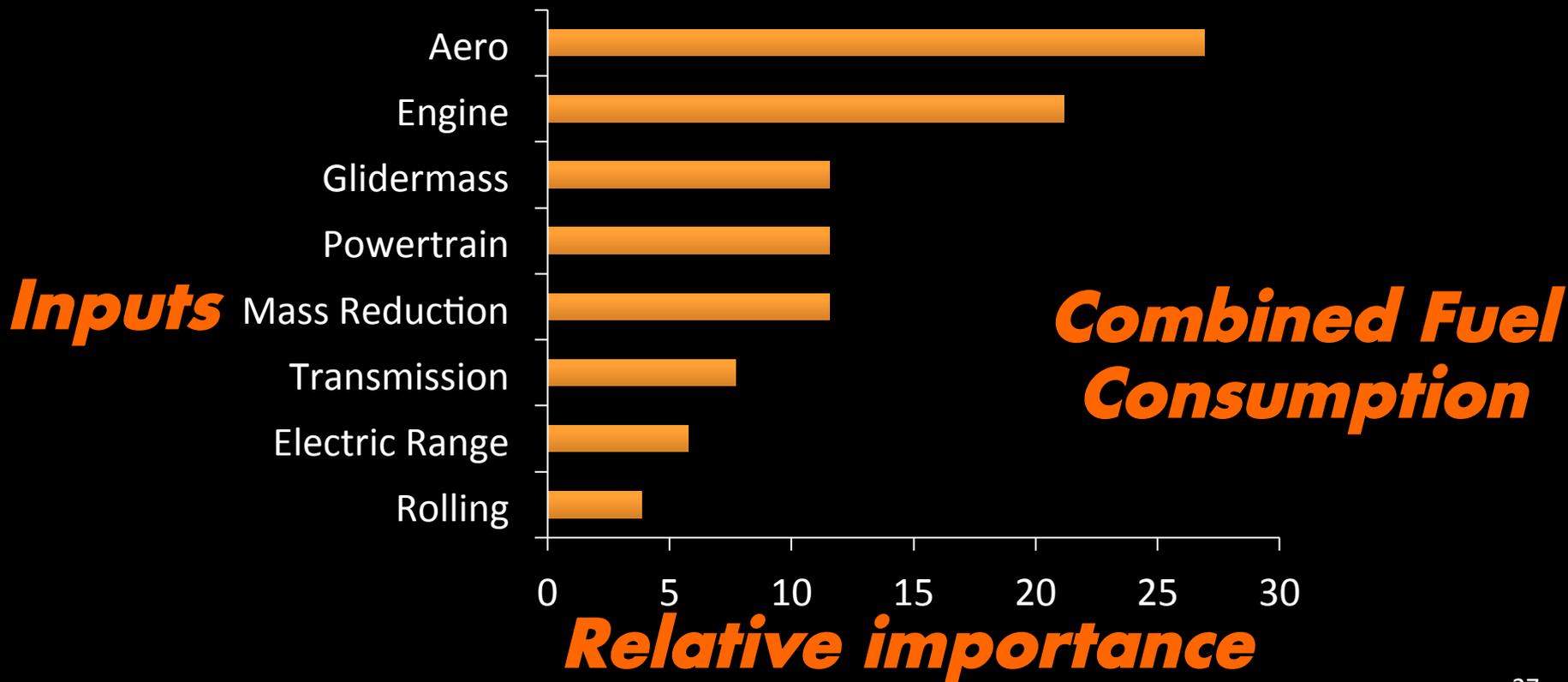
Simulation campaign reduction



Simulation campaign reduction



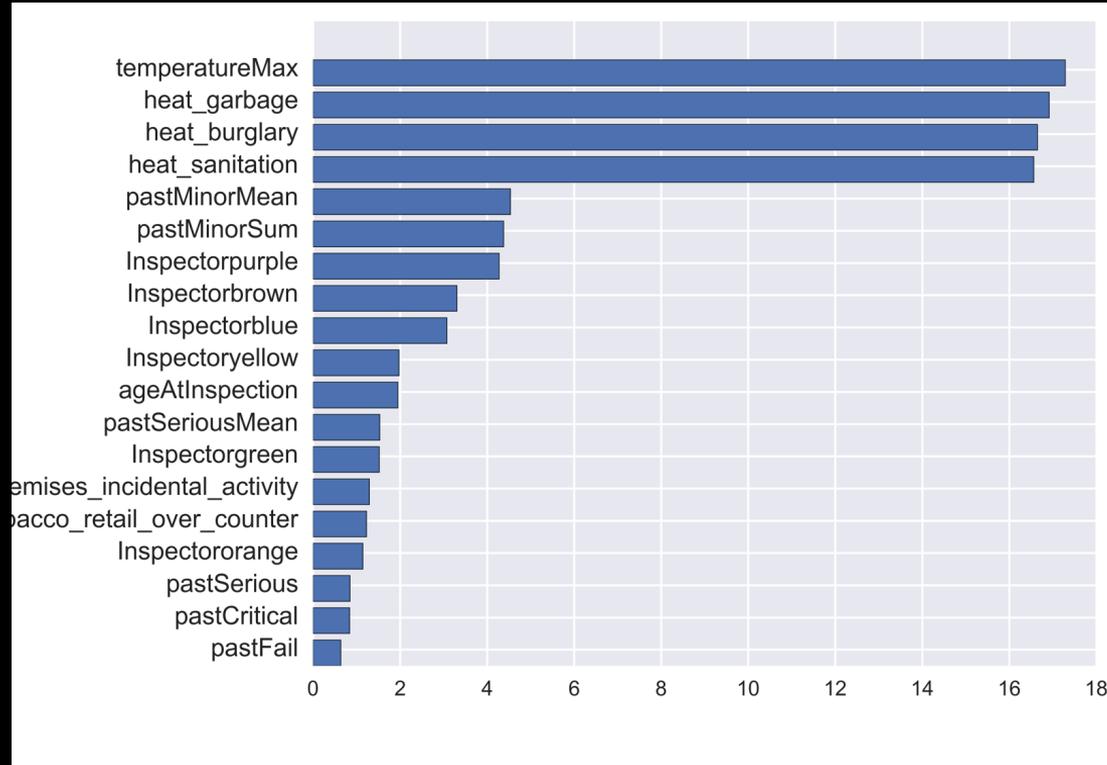
Simulation campaign reduction



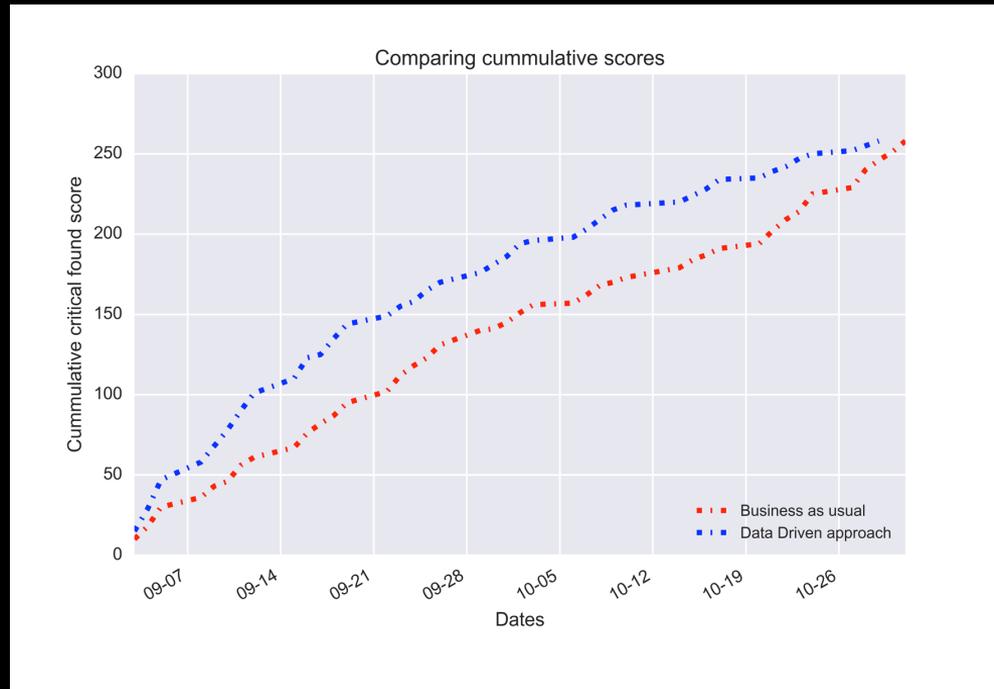
Restaurant violation in Chicago

Data Set Name	Data Set Size	Number of Variables
<i>Business Licenses</i>	470994	31
<i>Food Inspections</i>	97432	16
<i>Crime</i>	27600	2
<i>Garbage Cart Complaints</i>	27600	2
<i>Sanitation Complaints</i>	27600	2
<i>Weather</i>	1307	5
<i>Sanitarian Information</i>	67497	7

Restaurant violation in Chicago



Restaurant violation in Chicago



Predict violations 7.8 days in advance

Outline

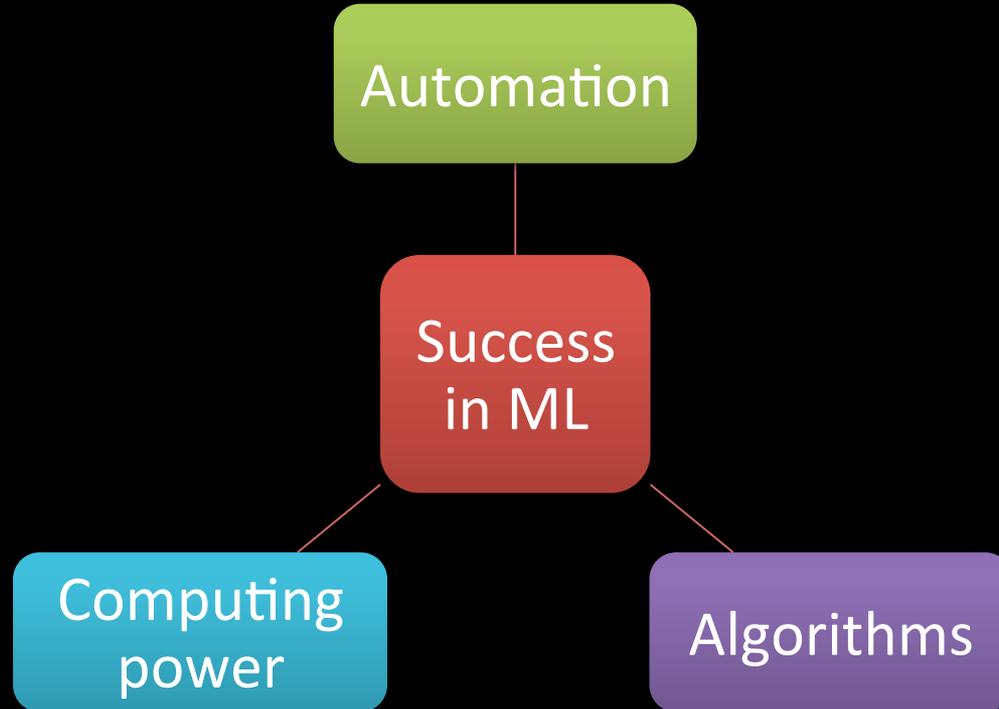
Motivation

AutoMOMML

Applications

Summary & Future

Summary



Thank You

www.mcs.anl.gov/~pbalapra