AutoML – Let the Machine do the Hard Work

Sascha Schubert

Senior Manager Artificial Intelligence

SAS

Who I am

linkedin.com/in/sschubertsas

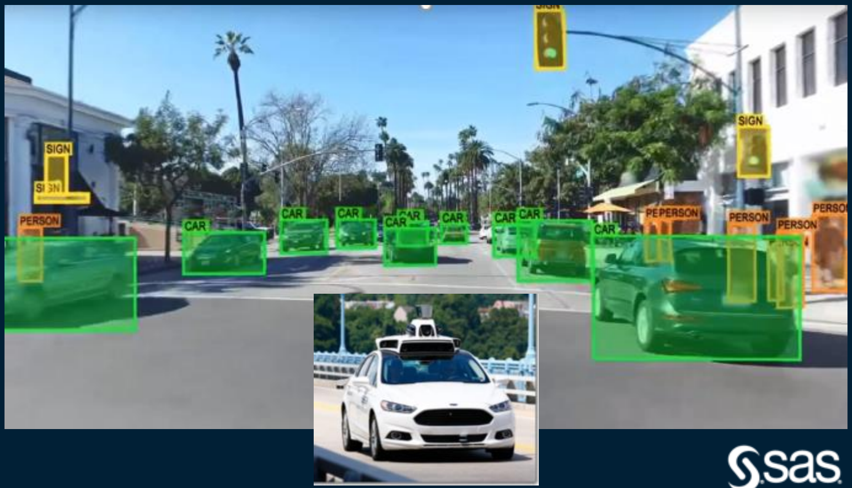
Agenda

1. Automation of Machine Learning
2. Key Steps of ML Pipelines
3. Automating ModelOps
4. Q&A


amazon
What can I help you with?


URGENT
38.57598
-121.4744
EncrVegt
SelMech
CAUTION
38.57523
-121.4768
EncrVegt
SelHerb


amazon go
JUST WALK OUT SHOPPING
amazon go


SIGN
PERSON
CAR

Automation

Traditional Process

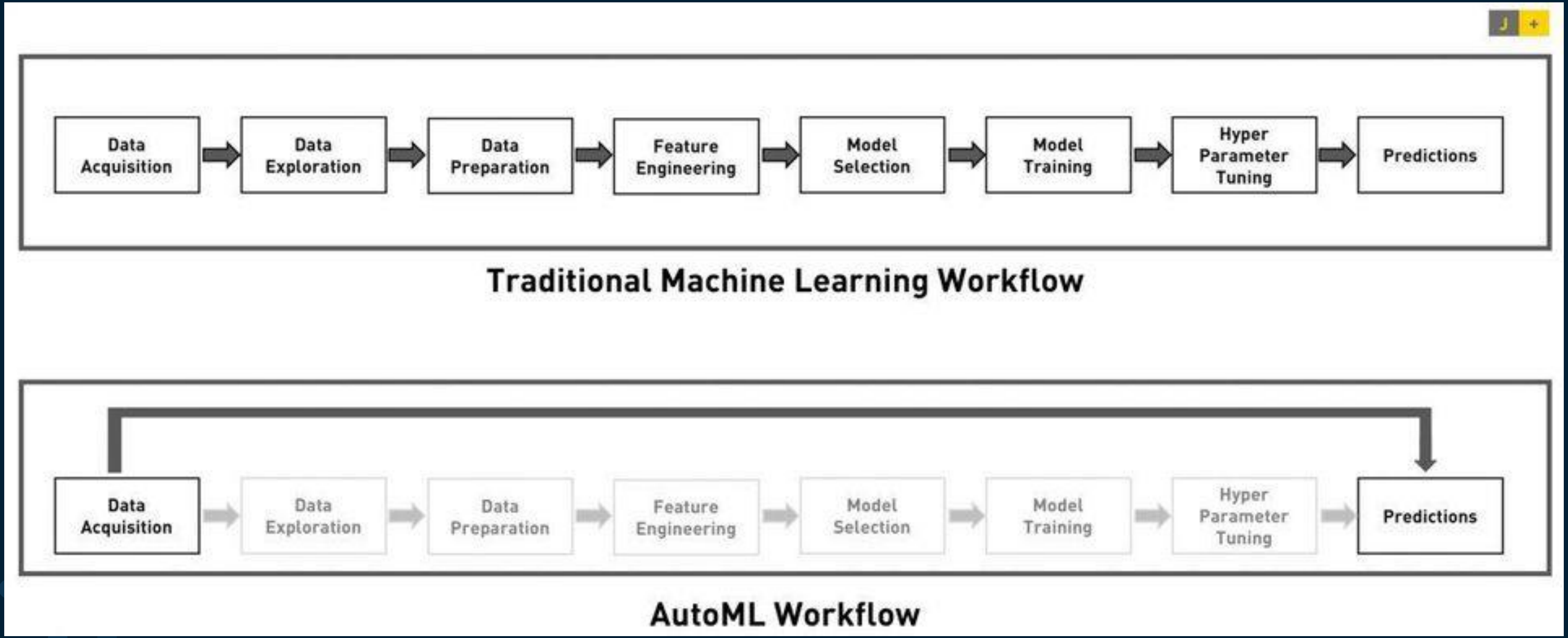
Data acquisition (sensors, etc)

AI/ML

Automated Process

Automated Machine Learning

Automating ML Pipelines



https://www.forbes.com/sites/janakirammsv/2018/04/15/why-automl-is-set-to-become-the-future-of-artificial-intelligence/#3c59a55f780a

Data Preprocessing

Generate High Quality Input Features

- Pre-Screen input to automatically determine the characteristics of variables and potential data quality issues
  - Class or numeric variable
  - Number of levels
  - Percentage of missing values

- Best Transformations
  - Identify best transformations for variables in relation to target
  - Automatically apply best transformation
  - Create new features

Automated Feature Generation

Generate High Quality Input Features

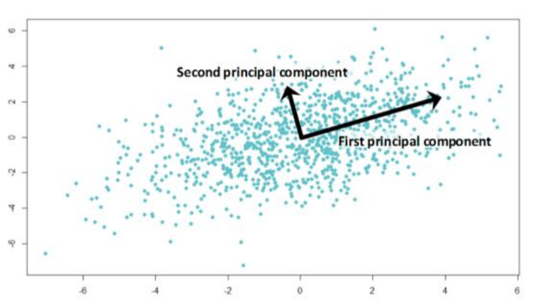
- Automatically create new features by performing variable transformations to fix data quality issues:
  - High cardinality – identify natural groups
  - High kurtosis – transform to maximize normality
  - High skewness - transform to maximize normality
  - Missing values – identify patterns of missingness
  - Outliers – identify and replace outliers
  - Collinearity – identify dependencies between features
  - Automatically detect interactions

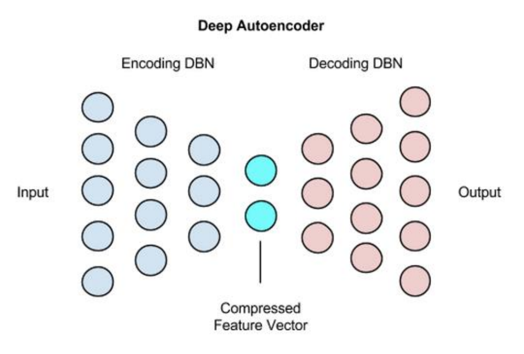
- Challenges:
  - **Input Variable Screening:** Define which features to transform based on data quality thresholds
  - **Feature Selection**: Define which new features to select based on feature – target relationship

Automated Feature Transformation

Combine Raw Variables

- Initial data set of raw features too large and correlated
- Automatically create features from raw data
- Features encapsulate the central properties of a data set and represent it in a low dimensional space
- Provide a more manageable, representative subset of input variables





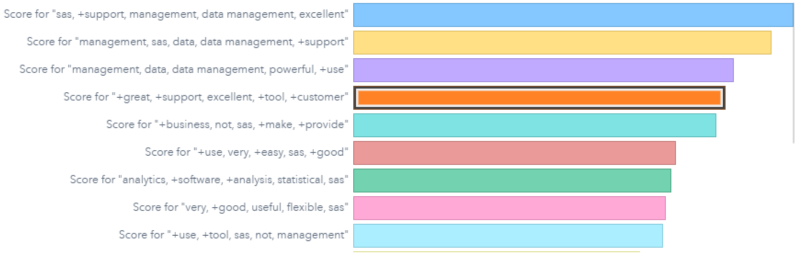
Automated Data Exploration

Promoter

What are the characteristics of Net Promoter Score in May 2019?

Net Promoter Score in May 2019 has a 8.65% chance (1.9K of 21K) of being Promoter. It's the second most common Net Promoter Score in May 2019 value.

What factors are most related to Net Promoter Score in May 2019?

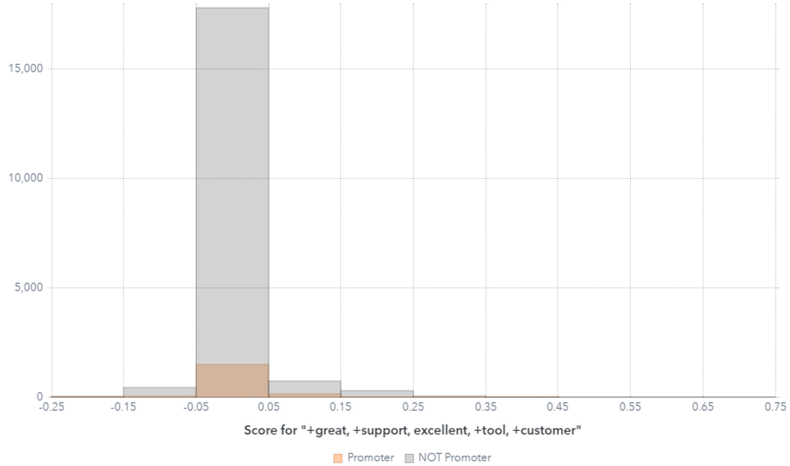


What are the groups based on Score for "+great, +support, excellent, +tool, +customer" by the chance of Net Promoter Score in May 2019 being Promoter?

High Low

96.15%	If Score for "+use, +tool, sas, not, management" is between .02 and .02, Score for "+great, +support, excellent, +tool, +customer" is less than -.04, then Net Promoter Score in May 2019 has a 96.15% chance (50 out of 52 cases) of being Promoter.
96.08%	If Score for "very, +good, useful, flexible, sas" is less than -0, Score for "+great, +support, excellent, +tool, +customer" is between -.04 and -.04, then Net Promoter Score in May 2019 has a 96.08% chance (49 out of 51 cases) of being Promoter.
83.10%	If Score for "+software, statistical, powerful, +statistical analysis, +support" is between -0 and 0, Score for "+great, +support, excellent, +tool, +customer" is between .02 and .02, then Net Promoter Score in May 2019 has a 83.10% chance (59 out of 71 cases) of being Promoter.

What is the relationship between Net Promoter Score in May 2019 and Score for "+great, +support, excellent, +tool, +customer"?

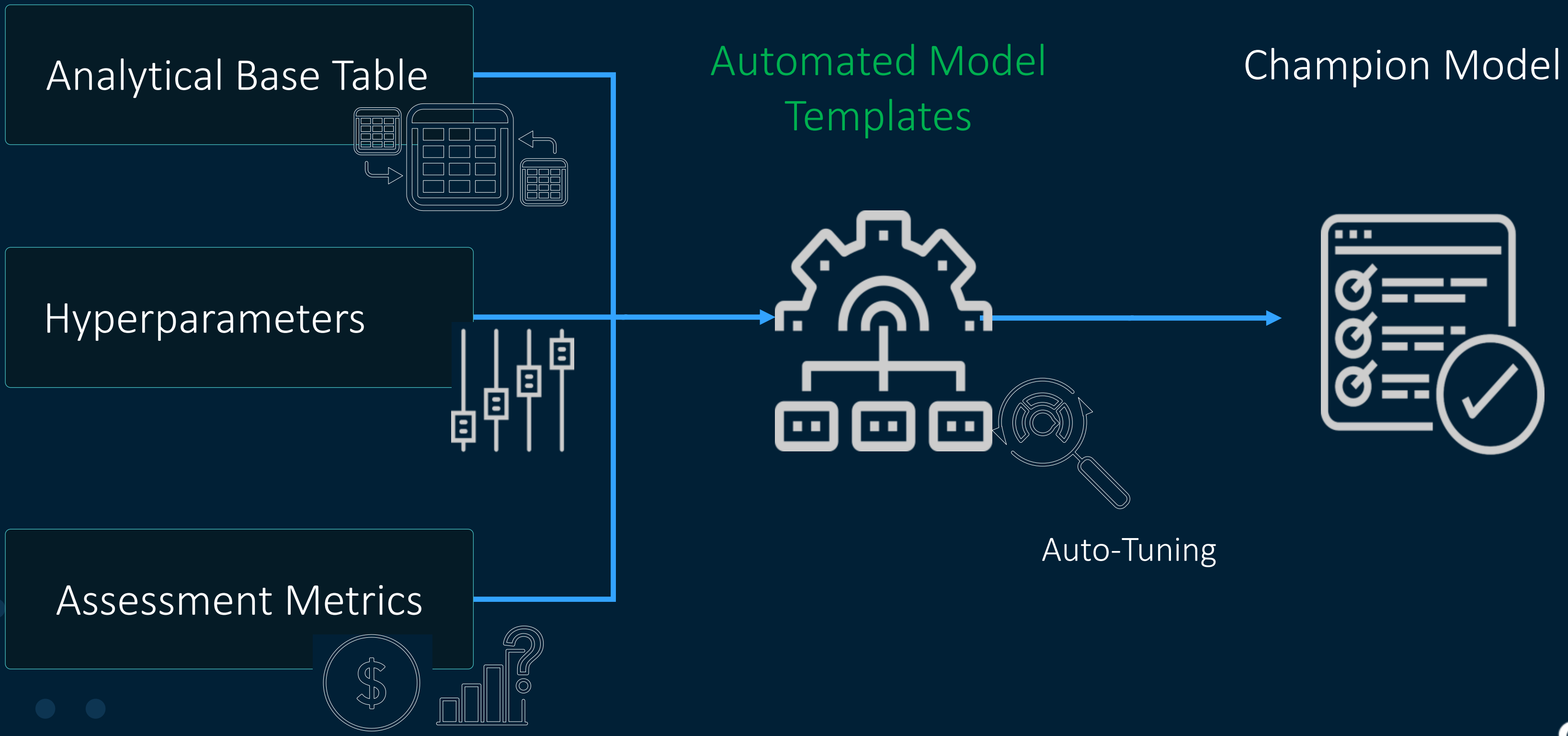


The average Score for "+great, +support, excellent, +tool, +customer" when Net Promoter Score in May 2019 is Promoter is .01, with a minimum of -.25 and a maximum of 0.4. The average Score for "+great, +support, excellent, +tool, +customer" when Net Promoter Score in May 2019 is NOT Promoter is .01, with a minimum of -.25 and a maximum of .71. Average Score for "+great, +support, excellent, +tool, +customer" is .01, and it ranges from -.25 to .71.

sas

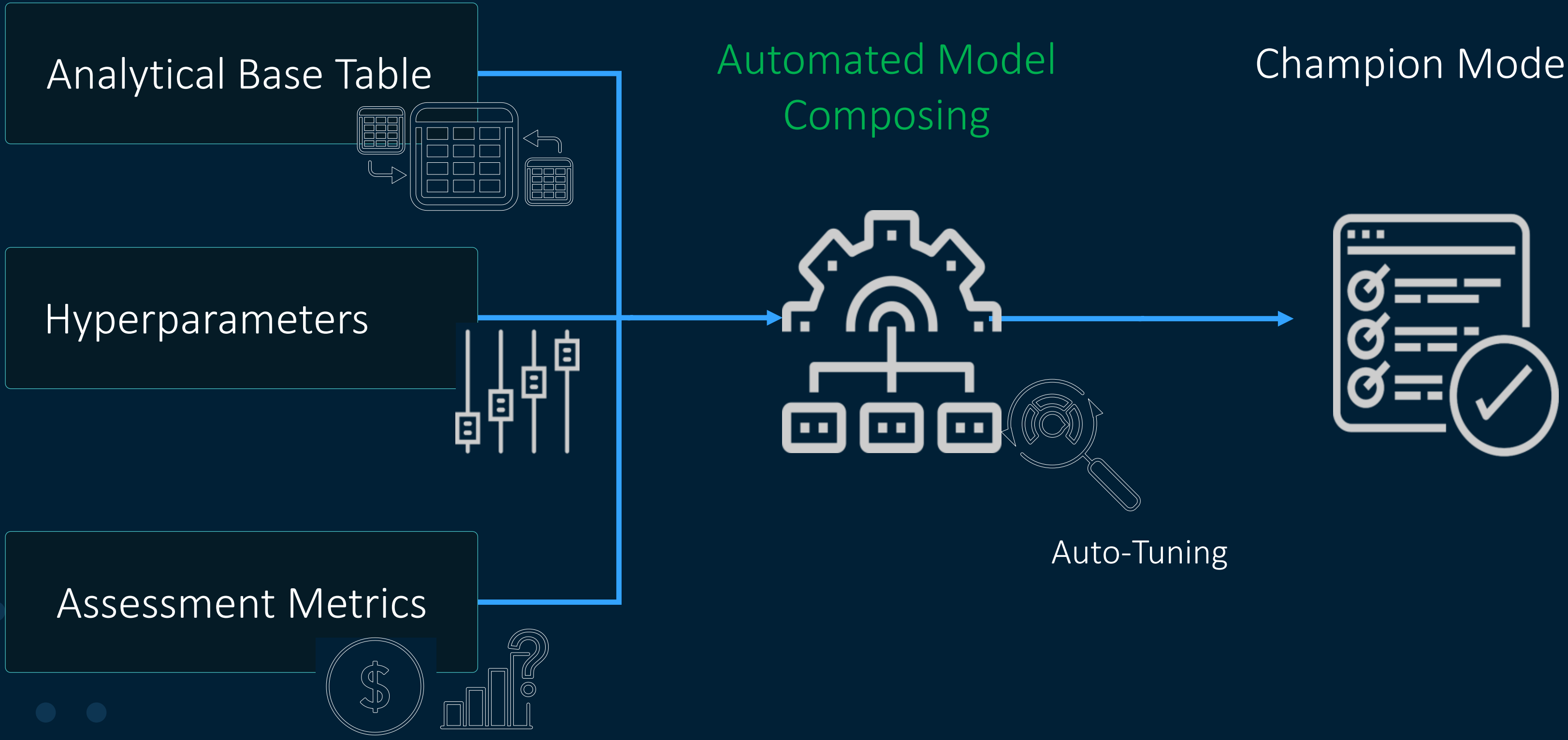
Automated Model Tournaments

Automated Static Templates



Automated Model Tournaments

Automated Dynamic Model Searches



Automated Hyperparameter Tuning

What are the hyper-parameters of a model?

- Training a model : determine **model parameters** or other logic to map inputs to a target
- Tuning a model : determine the **algorithm hyperparameters** (tuning options) that result in the model which maximizes predictability on an independent data set

Algorithm

Tuning options = hyperparameters

Model

- Regression coefficients
- Neural net weights
- Tree splitting rules
- …

Input

Target

Decision Trees

Max tree depth, splitting criterion, …

Neural Networks

Network architecture, solver options, …

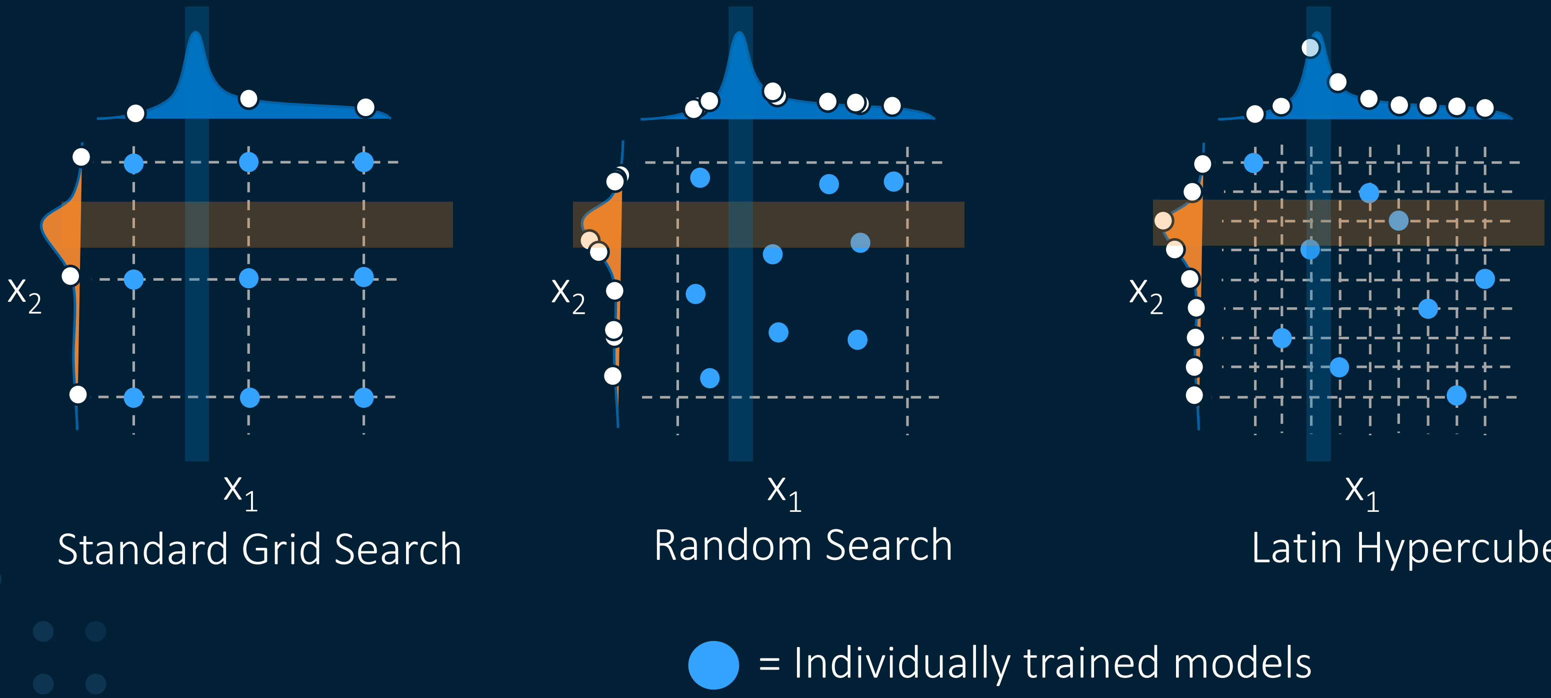
Support Vector Machines

Kernel, polynomial order, penalty, …

Very data/problem dependent!

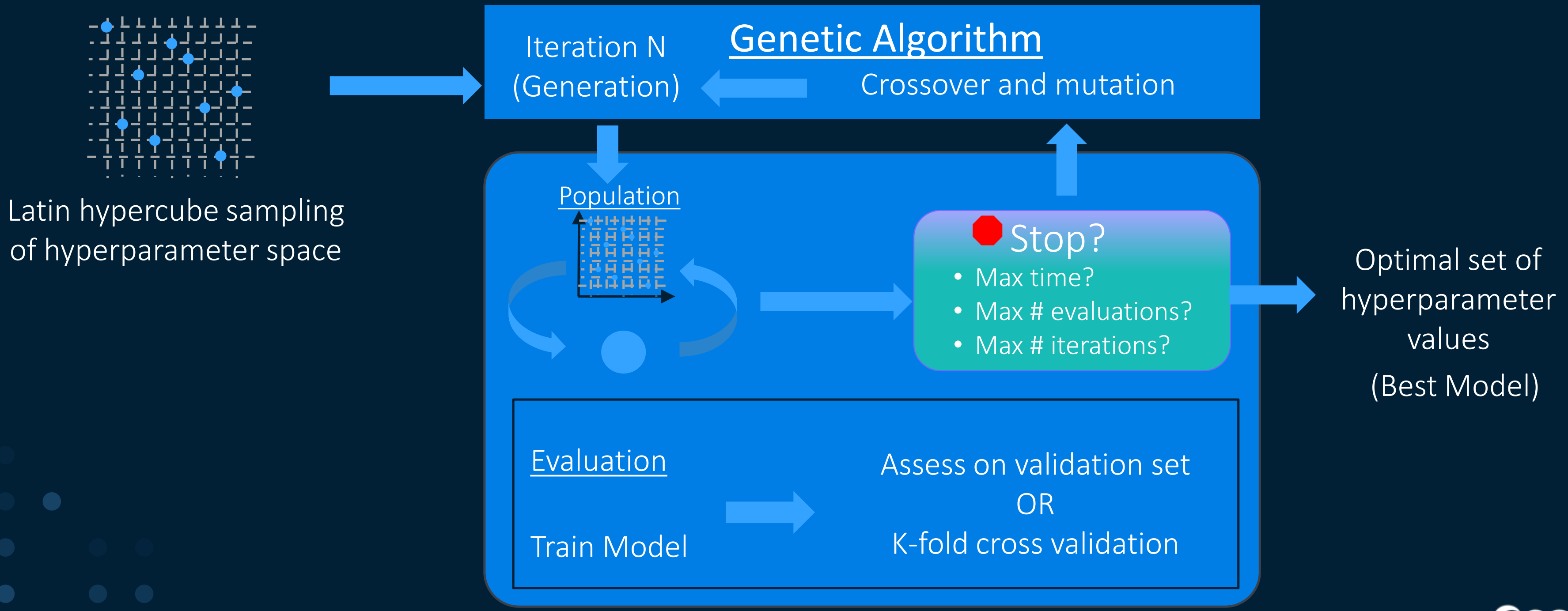
Hyperparameter Tuning – Typical Approaches

$y = f(x_1) + g(x_2)$



Automated Optimized Hyperparameter Tuning

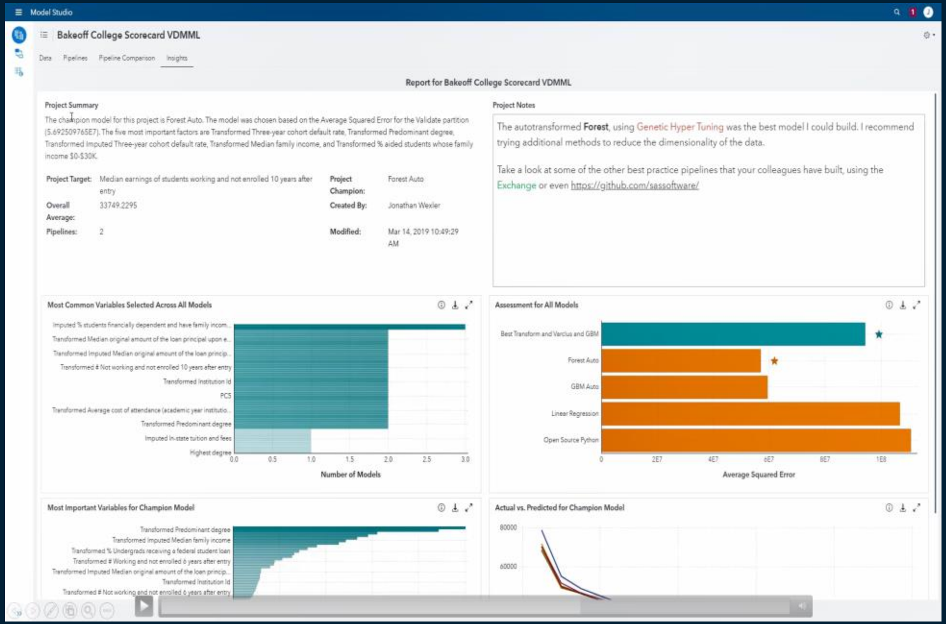
Formal optimization methods to intelligently search the hyperparameter space to find combinations which minimizes generalization error.



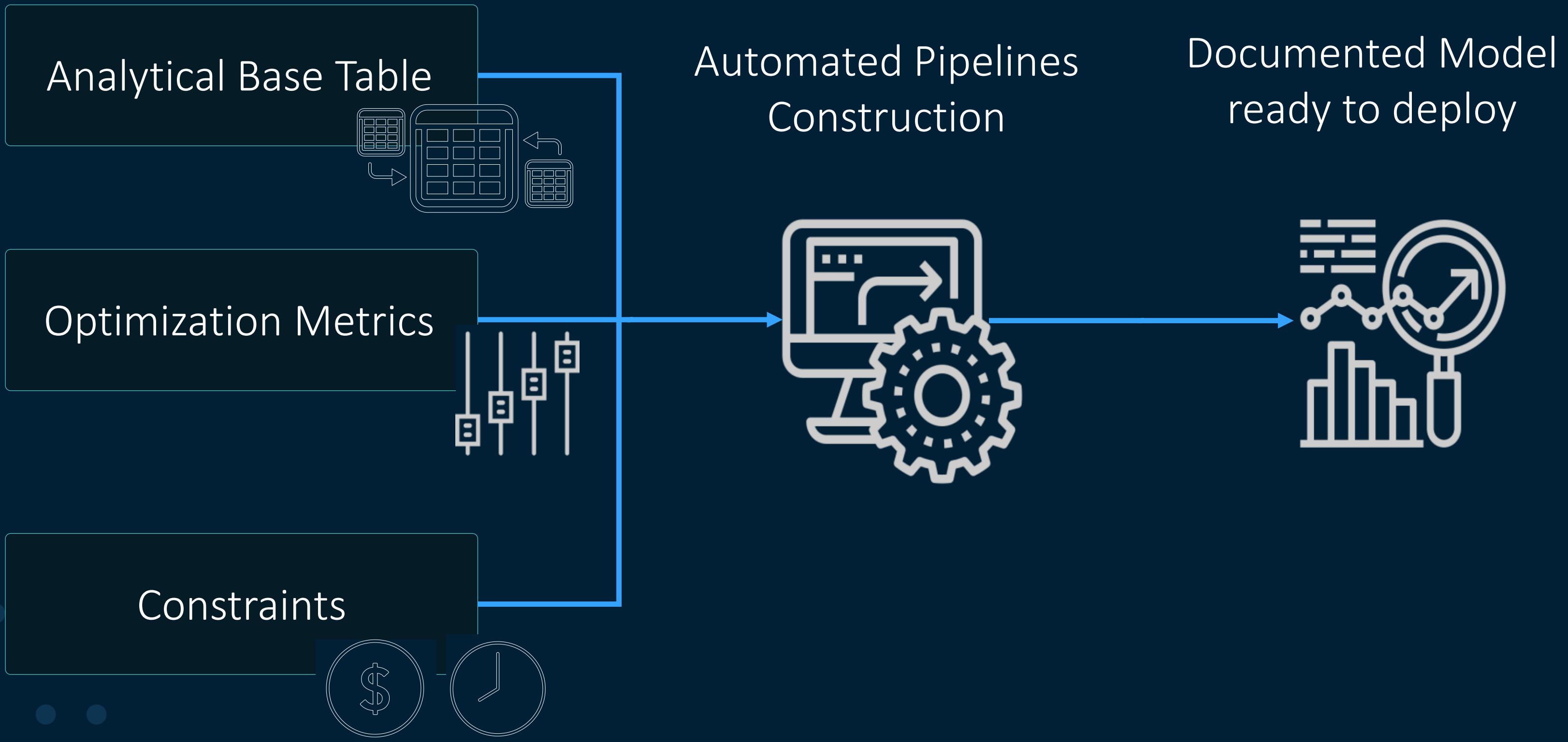
Automated Model Documentation

Modeling Insights

- Summarize model development project
- Report with graphs and description of results in plain English
- Use of NLG to generate reports based on model assessment statistics from model tournament



Automated Machine Learning Pipelines



DCSH COUNTY SOCIAL SERVICES
Automated Discovery Service

PROJECT NAME:

Child Safety Model

DATA SOURCE:

Select a table

GOAL:

Select a target

BUILD MODEL

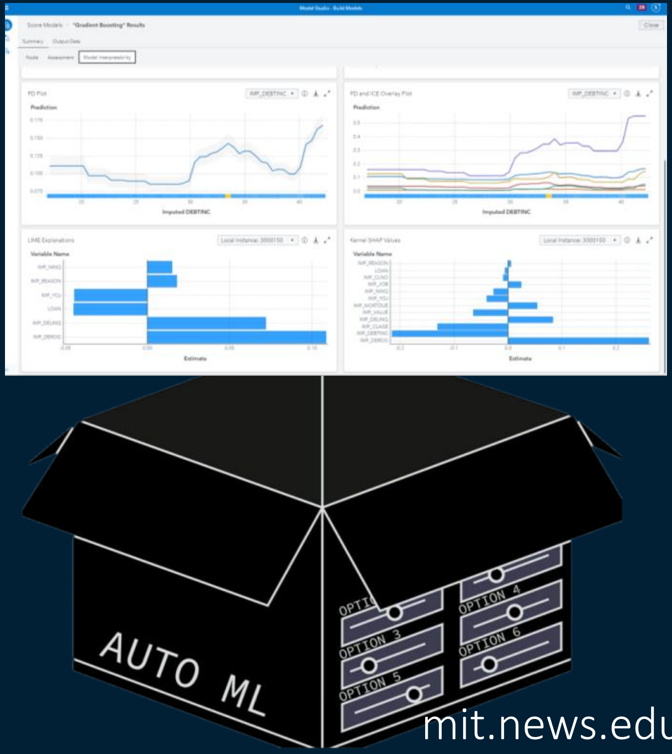
Analytics powered by SAS®

Set up your project and use the Build Model button to generate results.

Automated Explainability

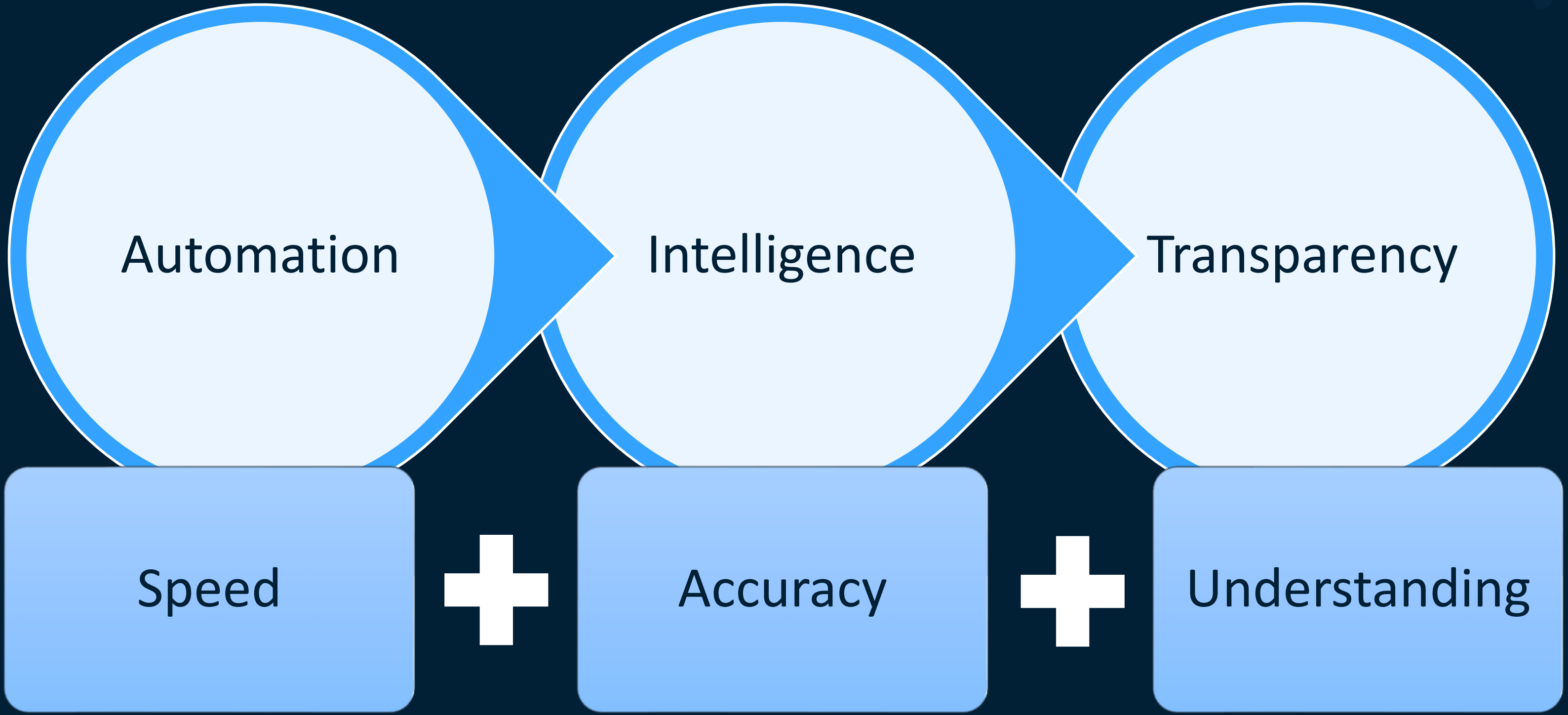
Make Black Box Transparent

- Make model understandable
- Explain model drivers
- Explain model stability
- Provide insights into single classifications
- Provide explanation of graphics in plain language



mit.news.edu

Formula for Success

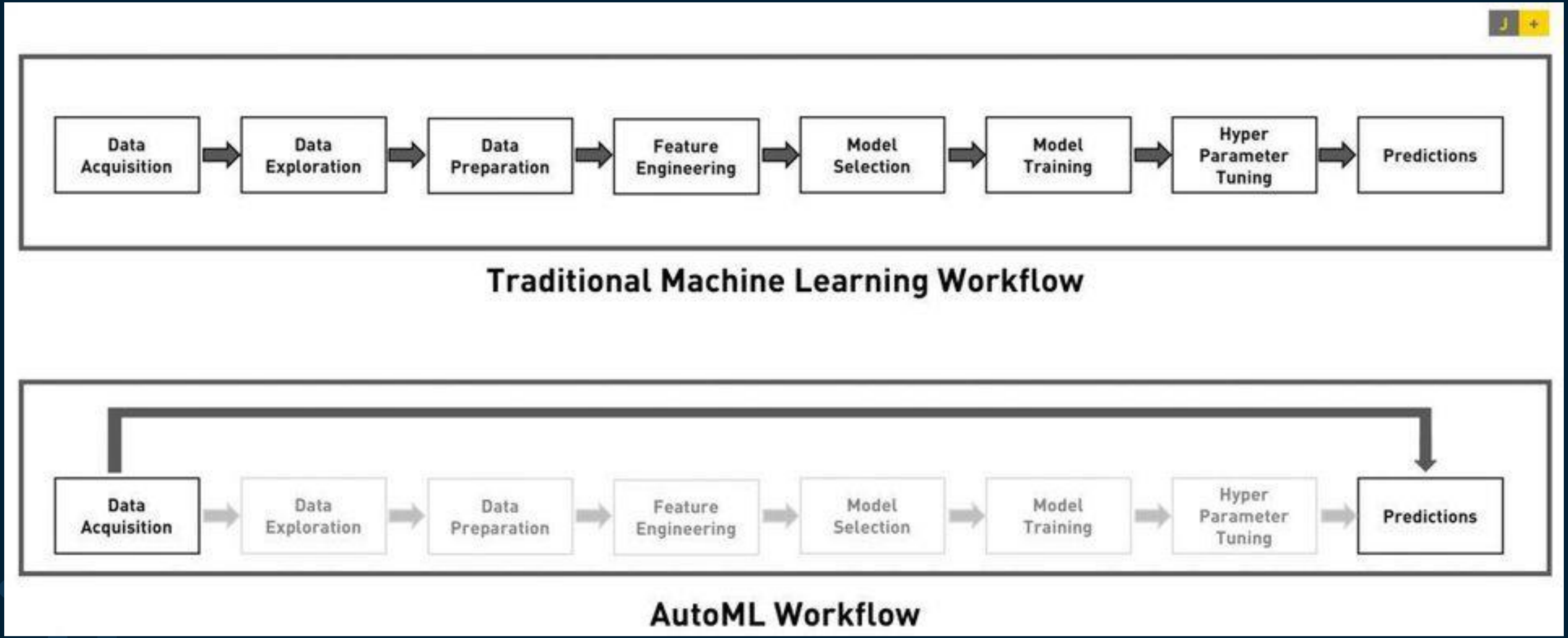


Where is AutoML Headed?

- Intelligence
  - Integrate AI into AutoML
- Explainability
  - White box
  - Ethical
  - Editable
- Flexibility
  - Cater to different levels of users with wizards, guidance, programming environments, APIs

Automated Machine Learning

Trusted Automated ML Pipelines

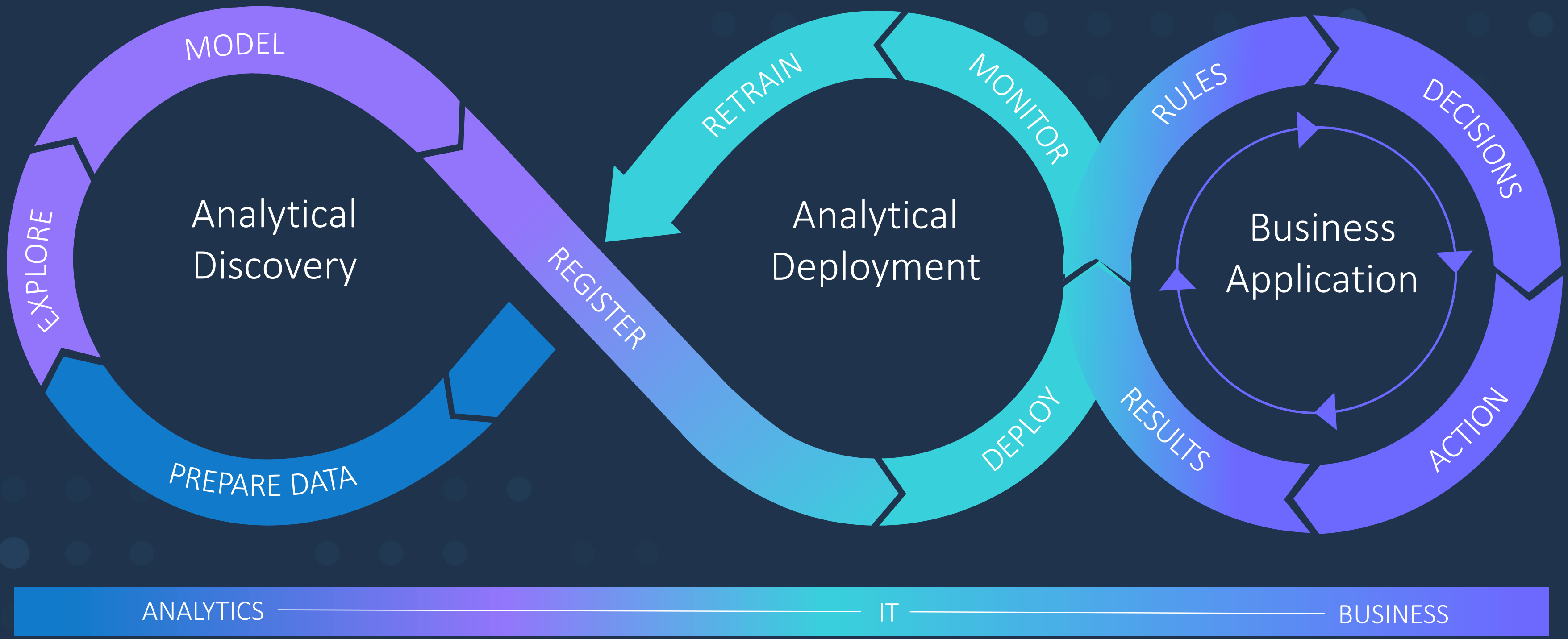


https://www.forbes.com/sites/janakirammsv/2018/04/15/why-automl-is-set-to-become-the-future-of-artificial-intelligence/#3c59a55f780a

Data doesn't change your organization, decisions do.

Every decision drives value, from big strategic choices to thousands of operational micro-moments.

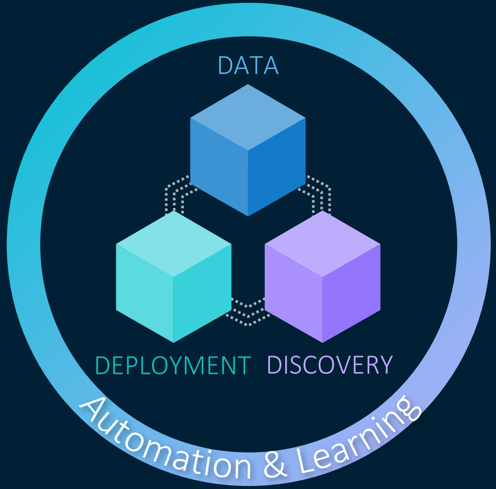
Analytical Decisioning Lifecycle



Intelligent Automation

Automation Hotspots

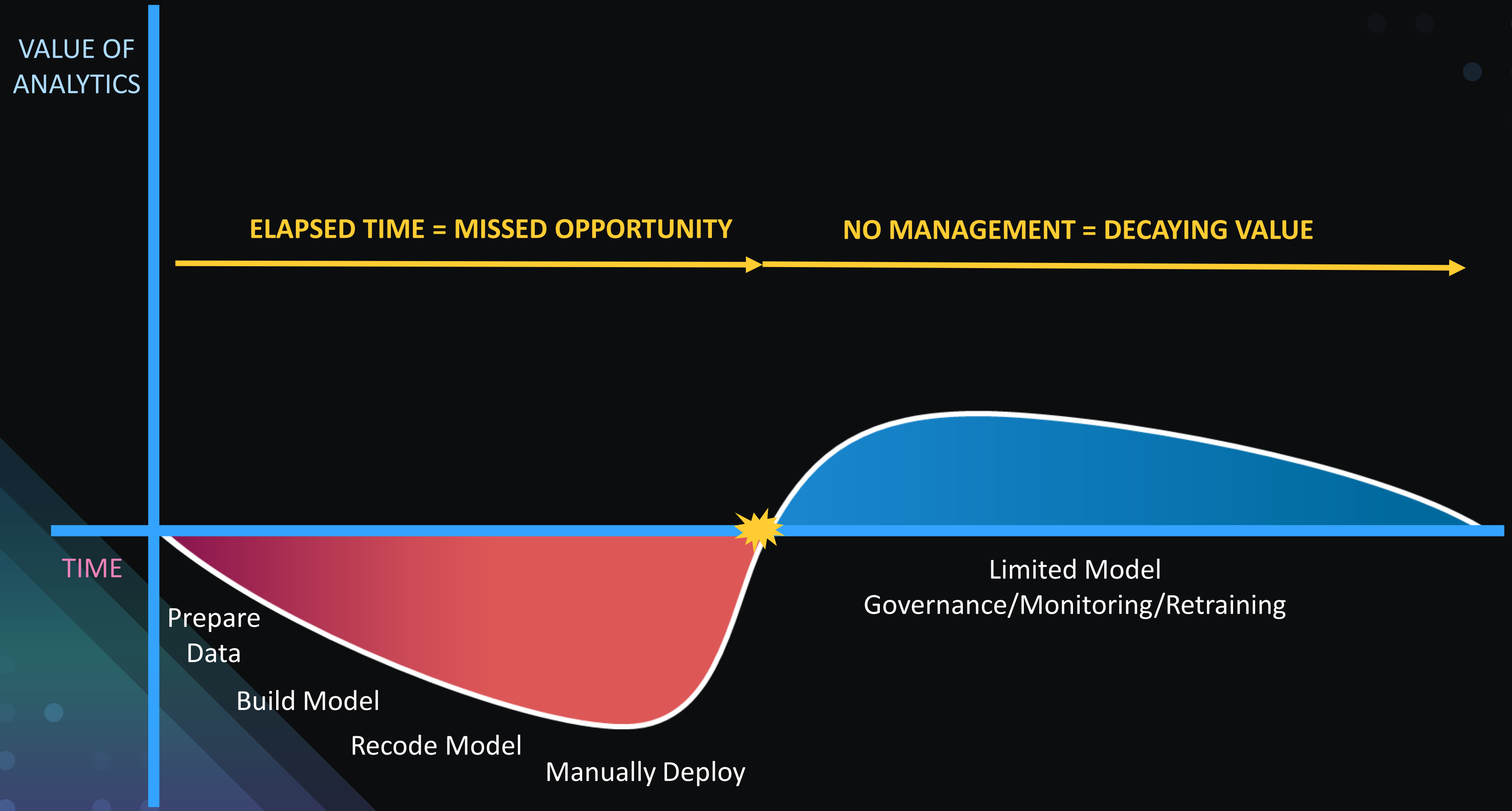
AI-based suggestions for data prep
Automated data cleaning and feature transformation



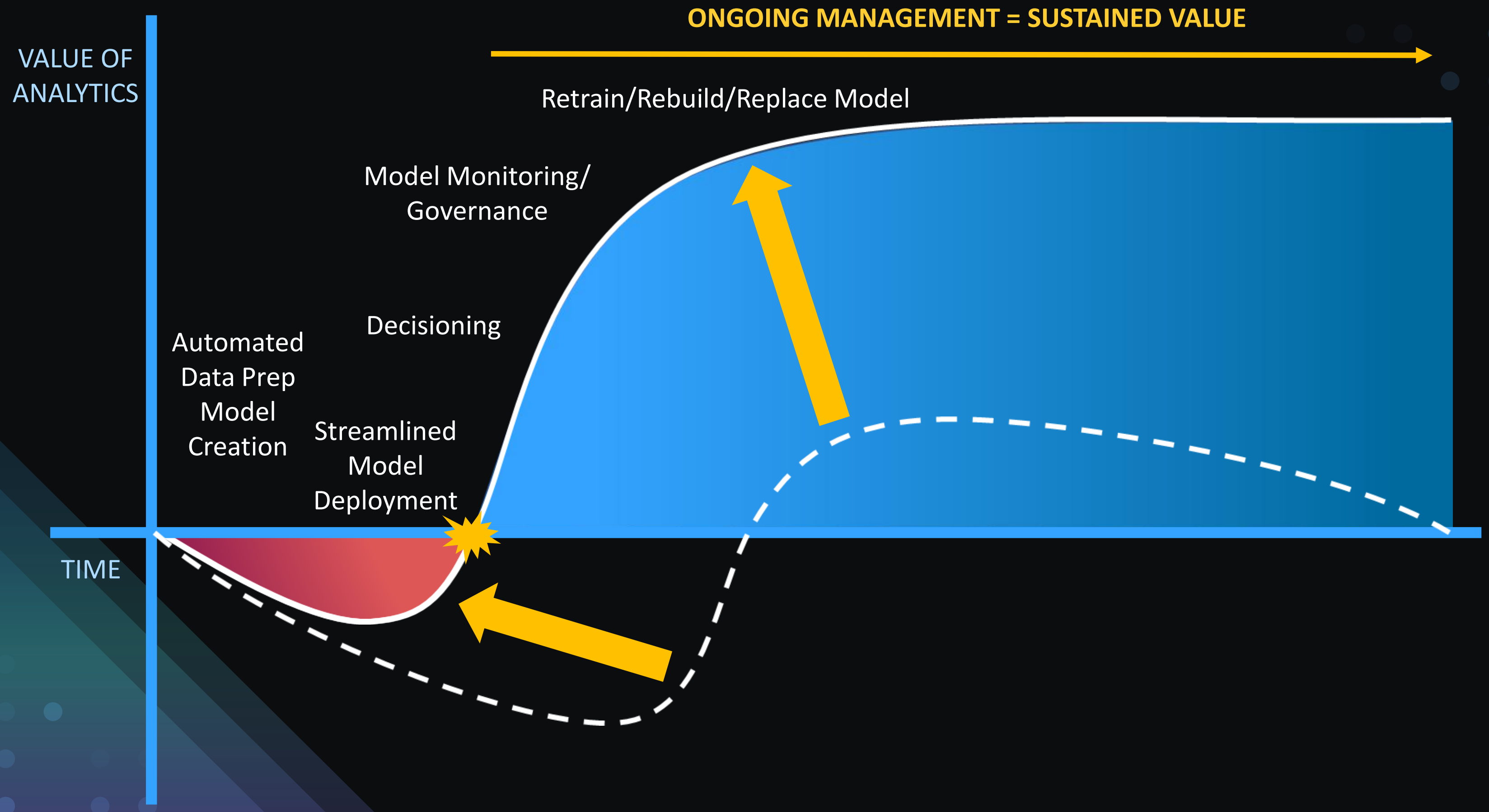
Automated model deployment
Automated model integration with decision flows
Automated model governance
Automated model retraining
Automated model lifecycle workflow

Automated feature generation
Automated model tuning
Automated model tournaments
Automated model optimization
Automated model assessment
Automated report generation

Accelerate Data to Decisions



VALUE OF
ANALYTICS
ELAPSED TIME = MISSED OPPORTUNITY
NO MANAGEMENT = DECAYING VALUE
TIME
Prepare
Data
Build Model
Recode Model
Manually Deploy
Limited Model
Governance/Monitoring/Retraining


ONGOING MANAGEMENT = SUSTAINED VALUE
VALUE OF ANALYTICS
Retrain/Rebuild/Replace Model
Model Monitoring/ Governance
Decisioning
Automated Data Prep Model Creation
Streamlined Model Deployment
TIME

ModelOps Defined

ModelOps is a framework that helps organisations to:

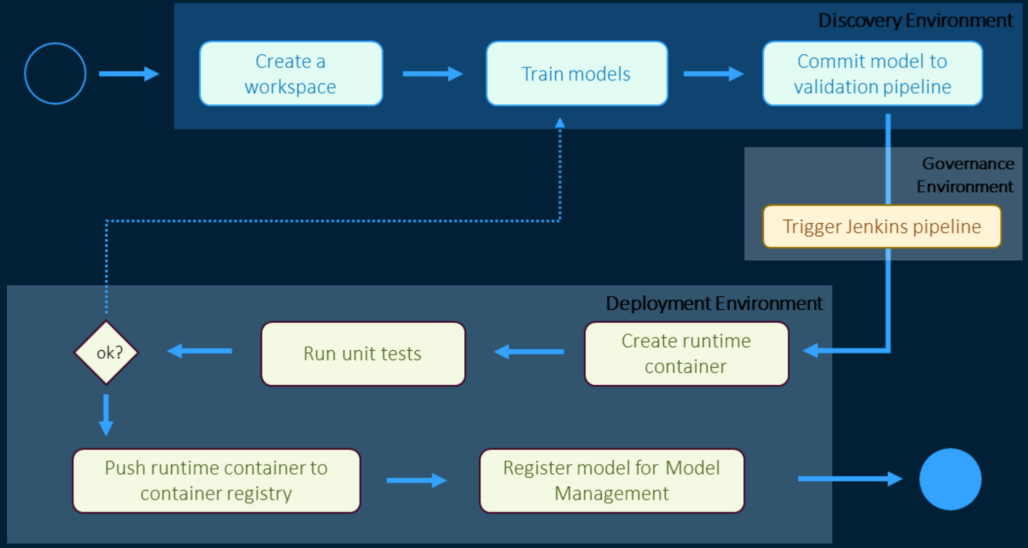
- Operationalize their analytics, i.e. to take models from development to production effectively, transforming their modelling efforts from an academic exercise to an economic benefit.
- Facilitate the collaboration between the different functions involved in managing analytic models as a corporate asset (data science, business units, IT & operations)
- Scale the use of predictive analytics for real-time decisioning

ModelOps Automation Through Pipelines

A Model Validation Pipeline

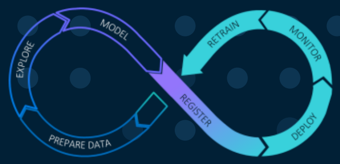
Example: Quality Assurance gateway for the model development process

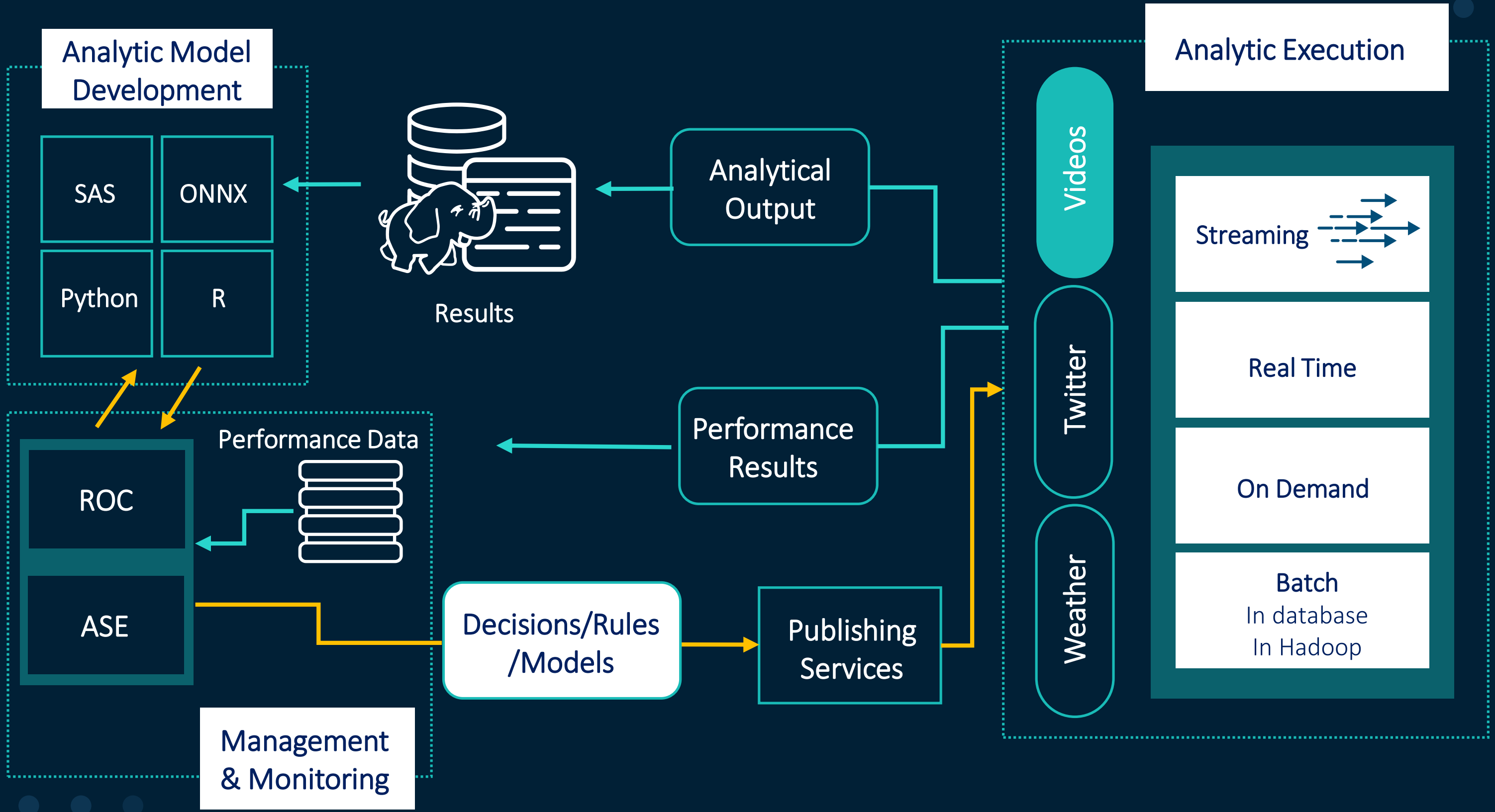
- Only models complying with your quality standards are allowed to be registered to the model repository
- A pipeline executing unit tests will be triggered whenever a data scientist tries to register a model



Automated Analytic Lifecycle

Dev – Test – Deploy – Monitor - Improve





More Information

- Blog Series: https://blogs.sas.com/content/tag/data-science-pilot-explained/
- SAS Visual Data Mining and Machine Learning, customized nodes examples: https://github.com/sassoftware/sas-viya-dmml-pipelines

- Documentation
  - SAS® Visual Data Mining and Machine Learning: https://support.sas.com/en/software/visual-data-mining-and-machine-learning-support.html
  - Developers Page: https://go.documentation.sas.com/?cdcId=pgmsascdc&cdcVersion=9.4_3.5&docsetId=casactml&docsetTarget=casactml_datasciencepilot_toc.htm&locale=en

Thank you!

Sascha.Schubert@sas.com

sas.com