

# SAS VDMML / SAS IML

## wprowadzenie i warsztaty

### dla UFG SA

[Łukasz.Chojak@sas.com](mailto:Łukasz.Chojak@sas.com)

[Mariusz.Dzieciatko@sas.com](mailto:Mariusz.Dzieciatko@sas.com)

[Piotr.Kaczynski@sas.com](mailto:Piotr.Kaczynski@sas.com)

[Mariusz.Zychowicz@sas.com](mailto:Mariusz.Zychowicz@sas.com)

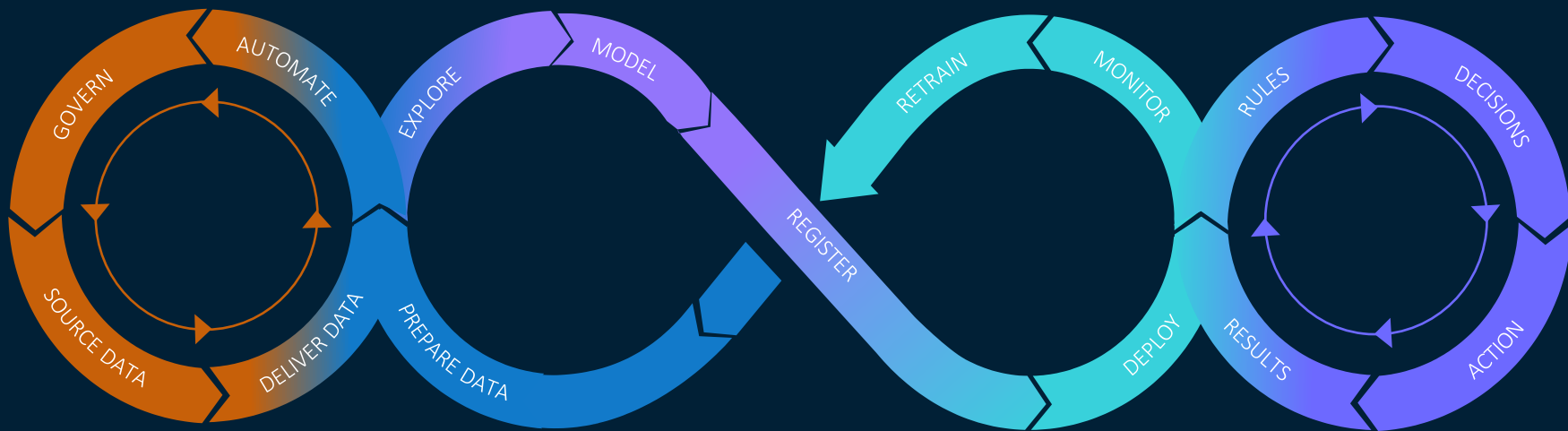
12.01.2021 Warszawa



# Agenda spotkania:

- Wprowadzenie do SAS VDMML i SAS IML
- Demo VDMML + demo koncepcja Open Modeling Platform (HUB dla zarządzania modelami SAS i open source).
- Dyskusja / Pytania

# Pełny cykl analityczny



Jeden interfejs

# SAS® IML

- Programowe podejście do operacji i analizy zorientowanych na macierz
- Tak jak PROC IML jest samotną procedurą w SAS/IML, akcja iml jest jedynym członkiem zestawu działań IML w SAS IML
- Za pomocą akcji iml można wykonać większość kodu PROC IML
- Możliwe jest przetwarzanie równoległe i rozproszone
- Interfejs:
  - Z poziomu SAS (SAS Studio)—PROC CAS wywołuje akcję iml
  - Języki Python, Lua , Java, R — Bezpośredni dostęp do akcji iml

# SAS<sup>®</sup> IML funkcjonalności

## High-level programming language

- Translate mathematical formulas into programs.
- Matrices and vectors are treated as high-level objects.
- Data structures such as lists and in-memory tables.
- Iterative loops, IF-THEN/ELSE logic, etc.
- More than 300 built-in functions.
- Define new functions and subroutines.
- In the iml action: Functions can execute in parallel on multiple threads.

## Matrix computations

- Read from a data table into a matrix.
- Multithreaded matrix operations: addition, multiplication, raise to power, etc.
- Solve linear systems.
- Matrix decompositions: Cholesky, eigenvalues, SVD, etc.
- Send matrices to functions.
- In the iml action: additional multithreaded matrix computations.

## Numerical analysis

- Linear algebra.
- Root-finding algorithms.
- Integrals and derivatives.
- Differential equations (DEs).
- Interpolation.
- In the iml action: Find multiple roots, solve multiple DEs, in parallel.

## Optimization

- Linear programming.
- Mixed-integer linear programming.

## Simulation

- Simulate from dozens of built-in distributions.
- Univariate: Normal, lognormal, exponential, etc.
- Multivariate: Normal, multinomial, Wishart, etc.
- Time series models: ARIMA and multivariate variations.
- Program custom simulations.
- Regression models.
- Spatial models.
- In the iml action: Generate independent samples in multiple threads.

## Statistics & data analysis

- Means and standard deviations.
- Correlation and covariance matrices.
- Tabulate categorical data.
- Implement hypothesis tests.
- Multivariate data analysis.
- In the iml action: bootstrap methods and permutation tests in parallel.

# SAS® Visual Data Mining and Machine Learning

Interfejs wizualny  
"drag & drop"



Interfejs  
programistyczny



Przygotowanie  
danych



Wizualna  
eksploracja



Modelowanie



Wdrożenie



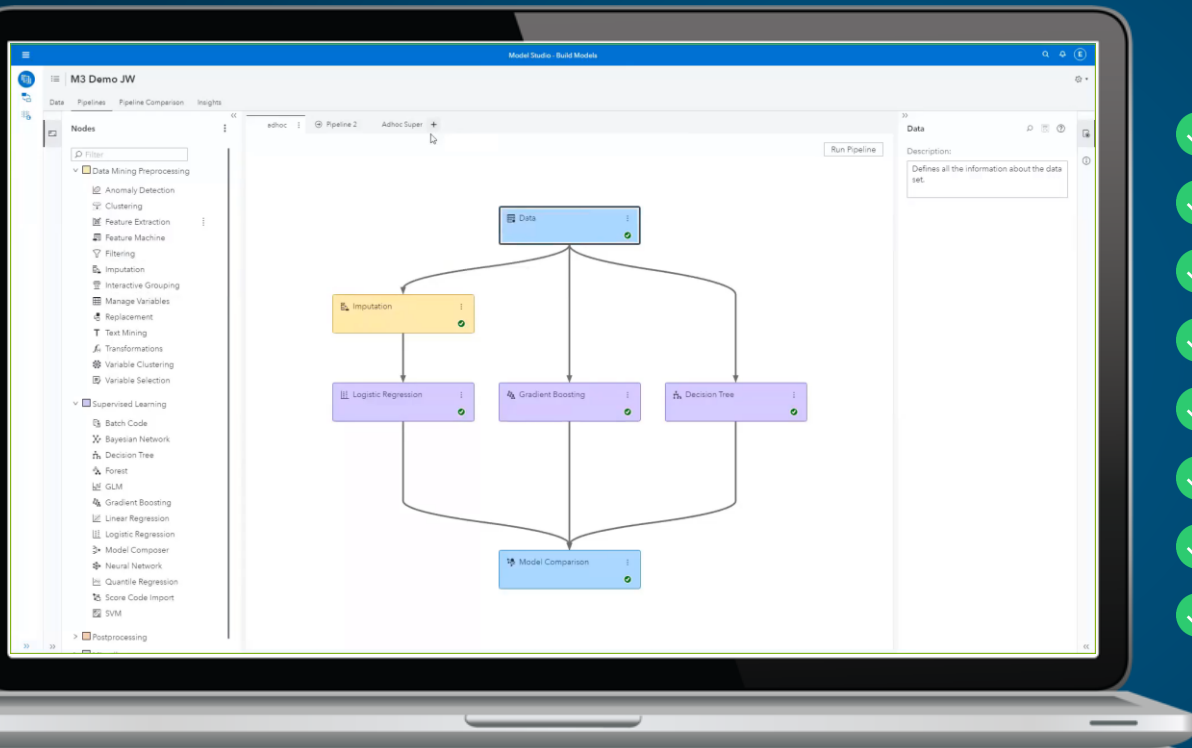
Uczenie  
maszynowe



Otwarte API



# Zautomatyzowane potoki



- ✓ Repository of best practice pipelines
- ✓ Models by SAS or by end-user
- ✓ Dynamically reads thru data
- ✓ Fixes data quality issues w/ ML
- ✓ Performs Data transformations
- ✓ Recommends & builds models
- ✓ Optimizes across models
- ✓ Fully editable, no black-box

# Automatyzacja za pośrednictwem interfejsu GUI lub wywołań interfejsu API

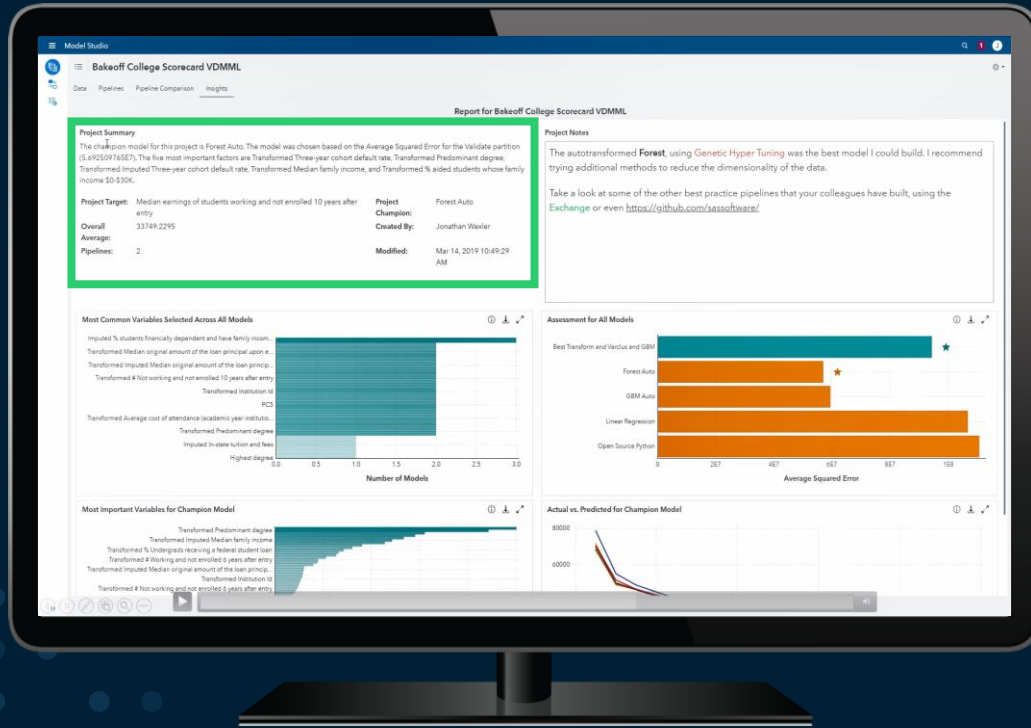
http://{host:port}/mlPipelineAutomation/projects

- ✓ Budowa modeli
- ✓ Retrenowanie modeli
- ✓ Skoring modeli
- ✓ Rejestracja modelu w repozytorium
- ✓ W pełni edytowalne

Dokumentacja do API SAS MLPA  
[developer.sas.com](https://developer.sas.com)



# Wytłumaczalność i możliwości NLP



- ✓ Model Interpretability Tab
- ✓ Explain models in plain language
- ✓ Project summaries using NLG
- ✓ PD – Partial Dependency
- ✓ LIME – Lcl. Interpret. Model-Ag. Expl.
- ✓ ICE – Individual Conditional Expl.
- ✓ SHAP – SHapley Additive exPlanations
- ✓ Collaboration w/ Project Notes

## Analityka

Analytic Astore Scoring  
Data Step  
Data Transpose  
DS2  
Function Compiler  
FedSQL  
Frequency / Crosstab  
Imputation  
Model Assessment  
Sampling and Partitioning  
Sentiment Analysis  
Sequencing/Pathing Analysis  
Text Mining  
Variable Binning  
Variable Cardinality Analysis  
Variable Clustering  
Variable Selection  
Variable Summary

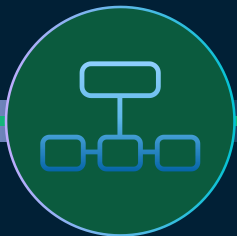
## Statystyka

Cox Proportional Hazards  
Decision Trees  
Design Matrix  
General Additive Models  
Generalized Linear Models  
Independent Component Analysis  
Clustering  
*K-means and K-modes*  
Linear Regression  
Linear Mixed Models  
Logistic Regression  
Model-Based Clustering  
Model Scoring  
Nonlinear Regression  
Ordinary Least Squares Regression  
Partial Least Squares Regression  
Pearson Correlation  
Principal Component Analysis  
*Kernel Principal Comp. Analysis*  
Quantile Regression  
Shewhart Control Chart Analysis

## Uczenie maszynowe

Automated ML  
*Data Science Pilot*  
*Model Composer*  
Audio Data Processing  
Bayesian Networks  
Boolean Rules  
Cross Validation ML  
Deep Learning  
*Convolutional NN*  
*Deep Forward NN*  
*Recurrent NN*  
*Transfer Learning*  
Factorization Machine  
Frequent Item Set Mining  
Gaussian Mixture Model  
Gaussian Process Regression  
Gradient Boosting  
Hyperparameter Auto-tuning  
K Nearest Neighbor  
Image Processing (*incl. Biomedical*)  
Market Basket Analysis  
Model Interpretability  
*LIME, ICE, PD, Shapley*  
Moving Windows PCA  
Multitask Learning  
Network Analytics  
Neural Networks  
Random and Isolation Forests  
Recommendation Engine  
Robust PCA  
Semi-supervised Learning  
Sparse Machine Learning  
Support Vector Data Description  
Support Vector Machine  
*t*-distributed SNE  
Text Parsing

# Computer Vision (CV)



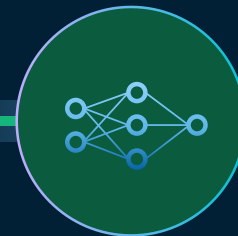
Nowe możliwości  
segmentacji



Wsparcie dla obrazów  
DICOM oraz  
DICOM RT



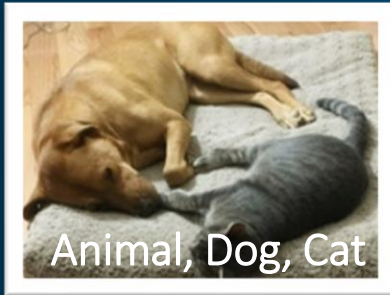
Poprawiona dokładność  
wykrywania obiektów–  
YOLOv2, TinyYOLO,  
R-CNN



CV na urządzenia  
mobilne–  
MobileNet,  
ShuffleNet

# Cztery zadania CV pokrywają ponad 90% przypadków użycia

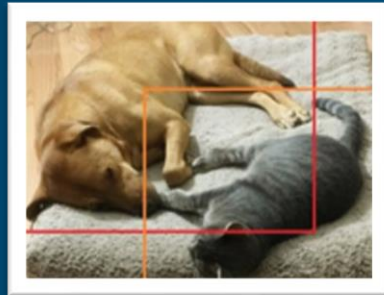
Klasyfikacja obrazów



**Zastosowanie:**  
Wykrywanie defektów na linii produkcyjnej

Convolutional Neural Network models for image classification such as VGG, ResNet, and Darknet, MobileNet, ShuffleNet

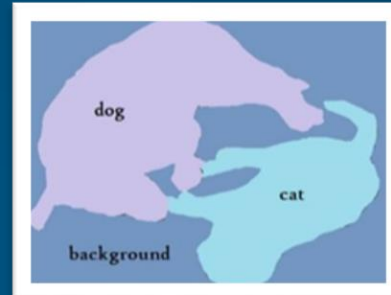
Detekcja obiektów



**Zastosowanie :**  
Monitorowanie zajętości miejsc parkingowych

YOLO-v1, YOLO-v2, Faster R-CNN

Segmentacja obiektów



**Zastosowanie :**  
Monitorowanie guza pod kątem zmiany kształtu, rozmiaru, koloru

UNET for semantic segmentation  
Mask R-CNN for instance segmentation

Detekcja kluczowych punktów

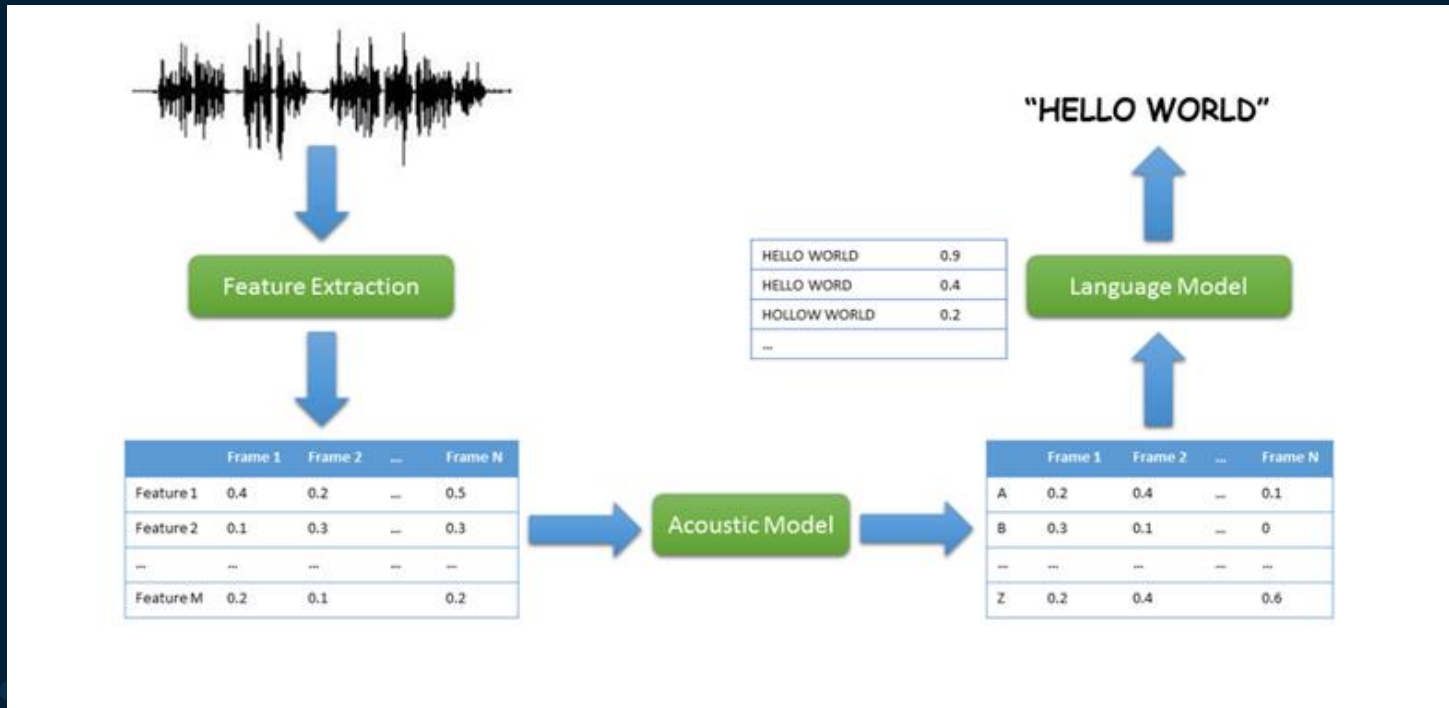


**Zastosowanie :**  
Ruch produktu na liniach produkcyjnych

Image regression & multi-task learning for keypoints detection

# Voice to text

## Architektura



# Nowe możliwości analityczne

VDMML vs. Miner

Nowe techniki  
analityczne

Automatyzacja i  
autotuning

Nowe typy  
danych  
(obraz i mowa)

Wydajność,  
skalowalność  
(in-memory)

Otwartość  
środowiska  
(OS + API)

# SAS® Visual Data Mining and Machine Learning

Otwartość dla nie SASowych programistów



Języki SAS

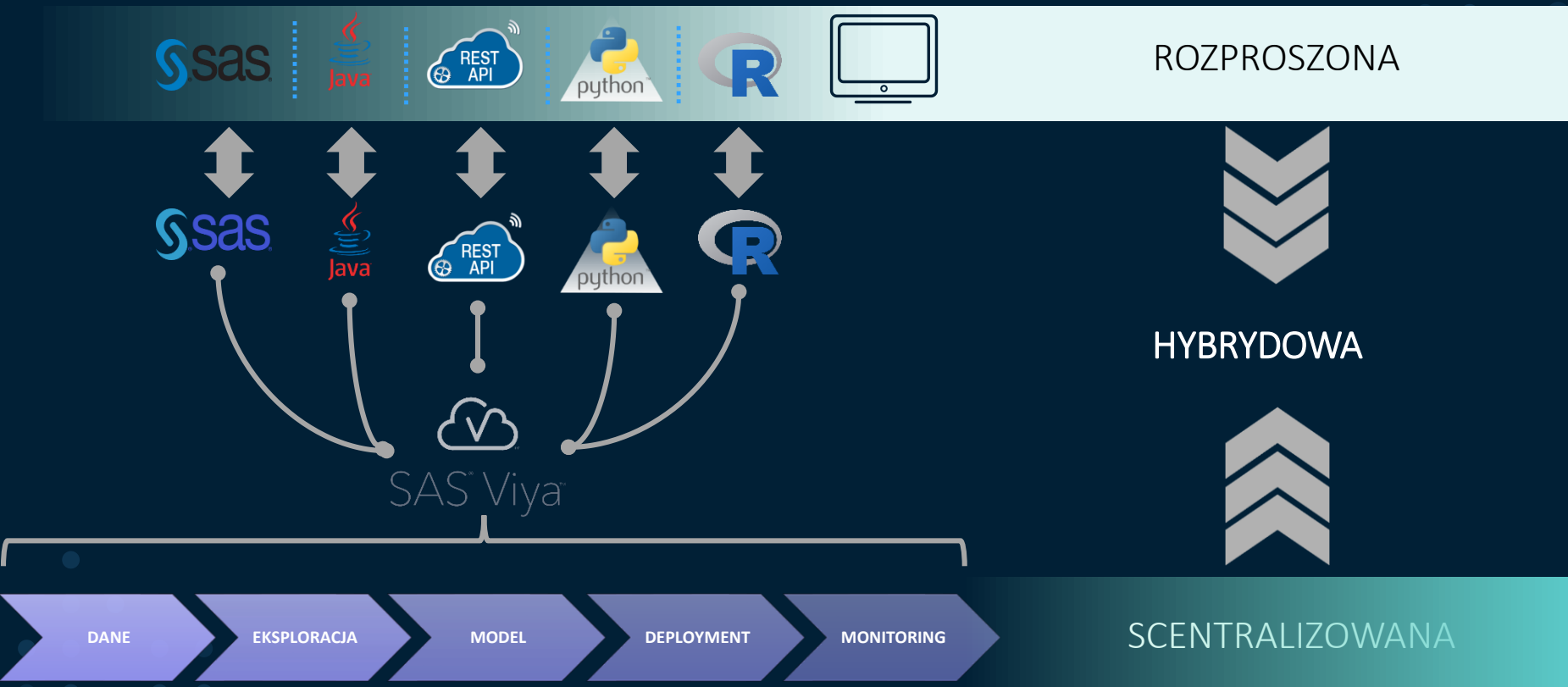


Python, R, Lua &  
Java



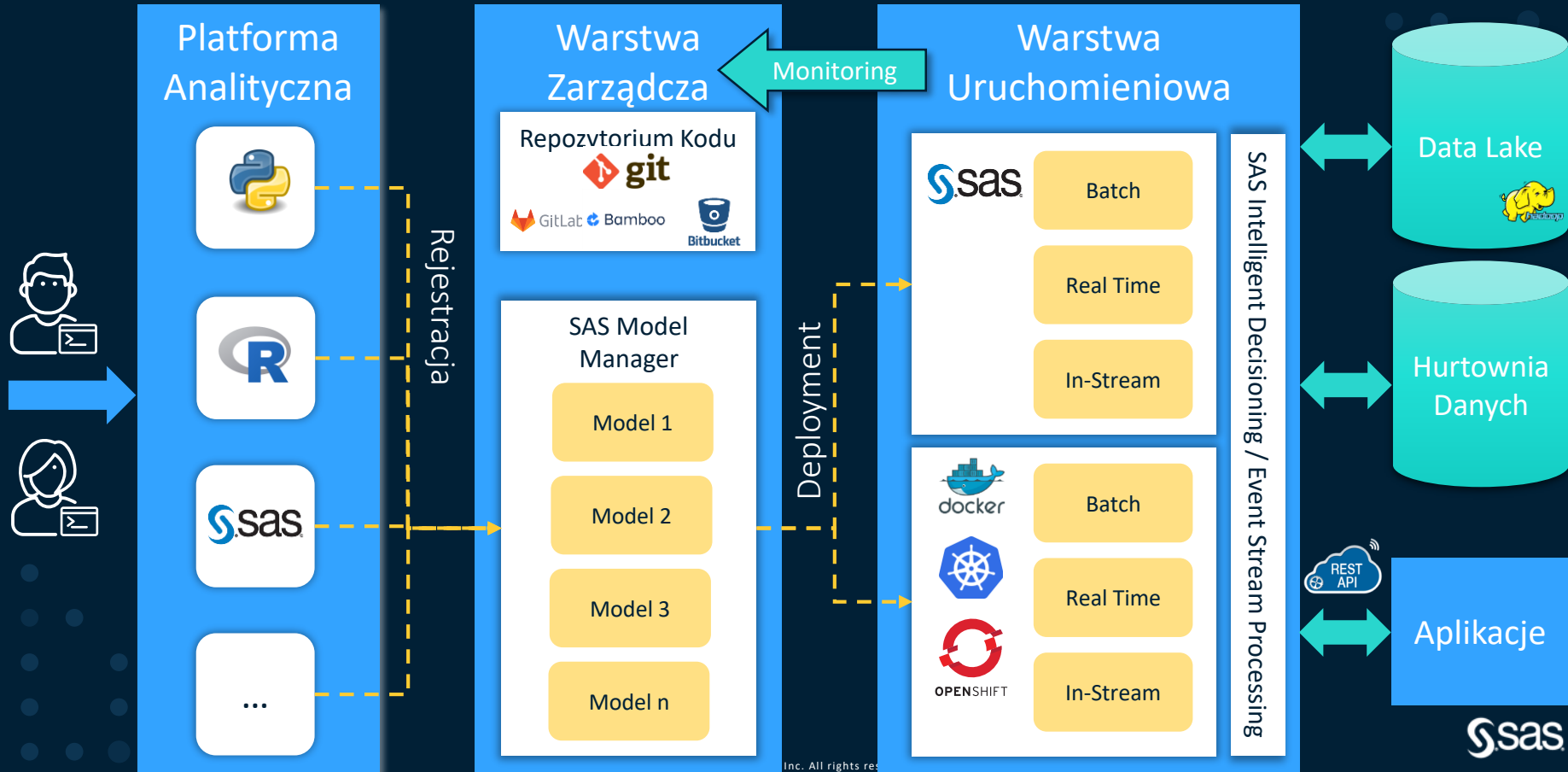
REST APIs

# Otwarty System Analityczny



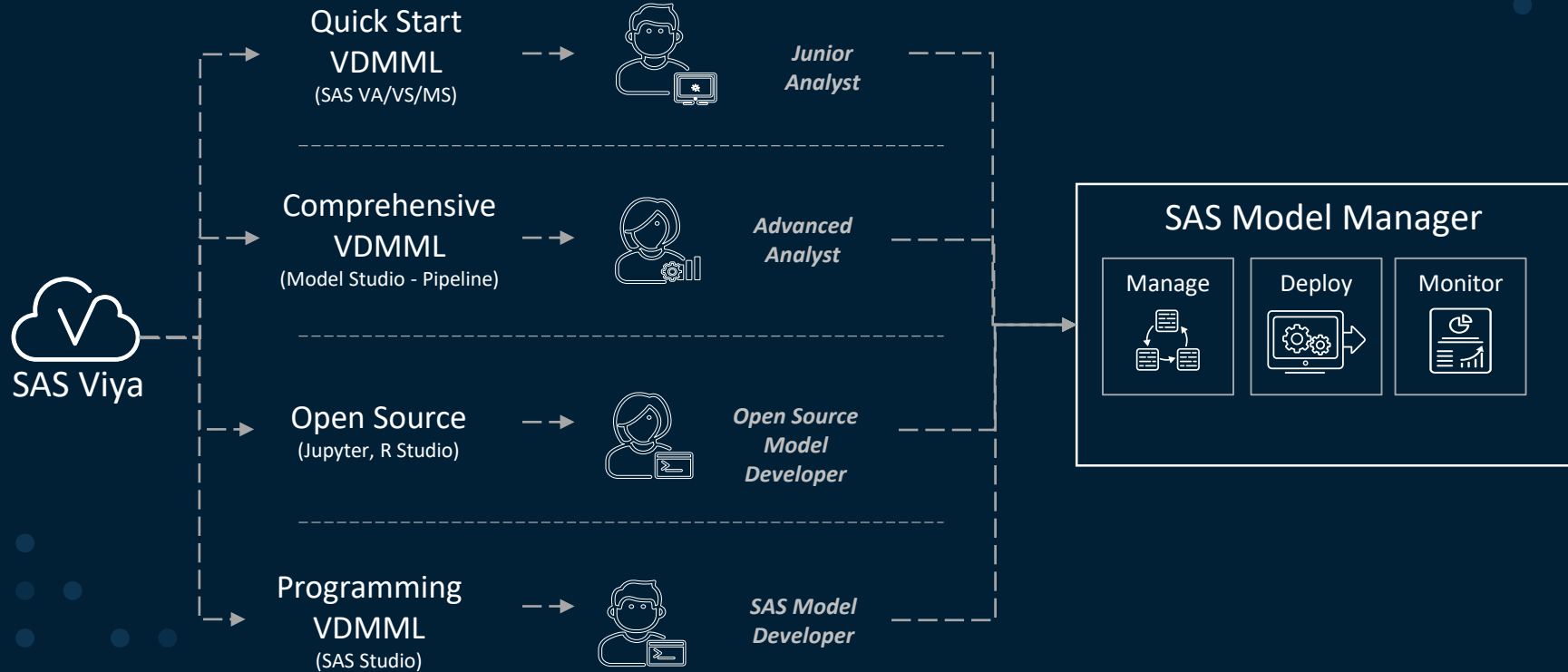


# Architektura logiczna



# Wykorzystanie SAS Viya

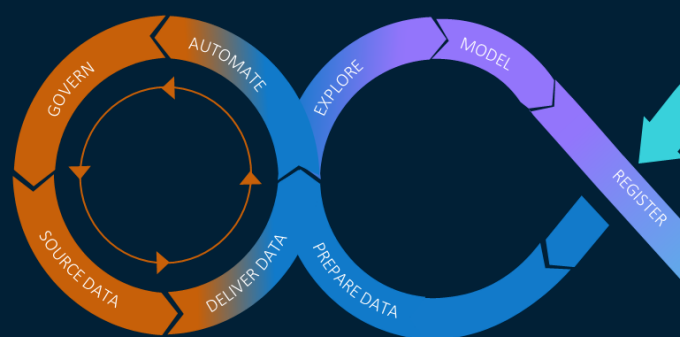
## Możliwe podejścia



DEVELOPMENT

DEPLOYMENT & MANAGEMENT





[DEMO]

## Tworzenie i orkiestracja modeli:

- VDMML – prezentacja narzędzia, cykl analityczny na modelach SASowych + integracja kodu OS.
- VDMML z wykorzystaniem interfejsu Python’a
- Tworzenie i rejestracja modeli OS w SAS Model Manager.

# Korzyści z wdrożenia Open Modeling Platform

## Perspektywa analityka



### Elastyczność

- Środowisko na żądanie
- Dowolna technologia
- Dostępność danych



### Skuteczność

- Techniki analityczne
- Autotuning
- Automatyzacja procesów



### Wygoda

- Wizualny interfejs
- Gotowe zadania
- Wydajne przetwarzanie

# Korzyści z wdrożenia Open Modeling Platform

## Perspektywa architekta/IT



### Bezpieczeństwo

- Dane w centralnym silniku
- Nadzór nad modelami
- Uprawnienia



### Skalowalność

- Przetwarzanie in-memory
- Architektura rozproszona
- Łatwość zarządzania



### Nowoczesność

- Konteneryzacja
- Best-practice (Git, Rest API, itd.)
- Gotowość na chmurę

# Korzyści z wdrożenia Open Modeling Platform

## Perspektywa kierownictwa



### Ograniczenie ryzyka

- Dane w jednym miejscu
- Rejestr modeli
- Nadzór nad cyklem analitycznym



### Wzrost efektywności

- Skuteczniejsze modele
- Krótki time-to-market
- Automatyzacja procesów



### Strategiczny kierunek

- Gotowość na chmurę



**Dziękujemy!**