

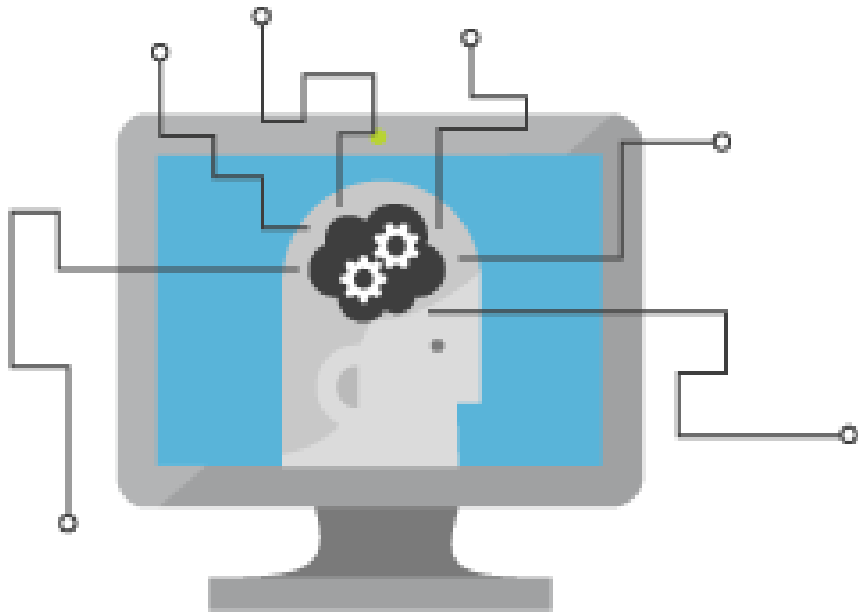
# (Azure) Machine Learning

Tomasz Kopacz, Microsoft

35 Minut + 5 pytania – taki jest plan!



# Co to Machine Learning (ML)?



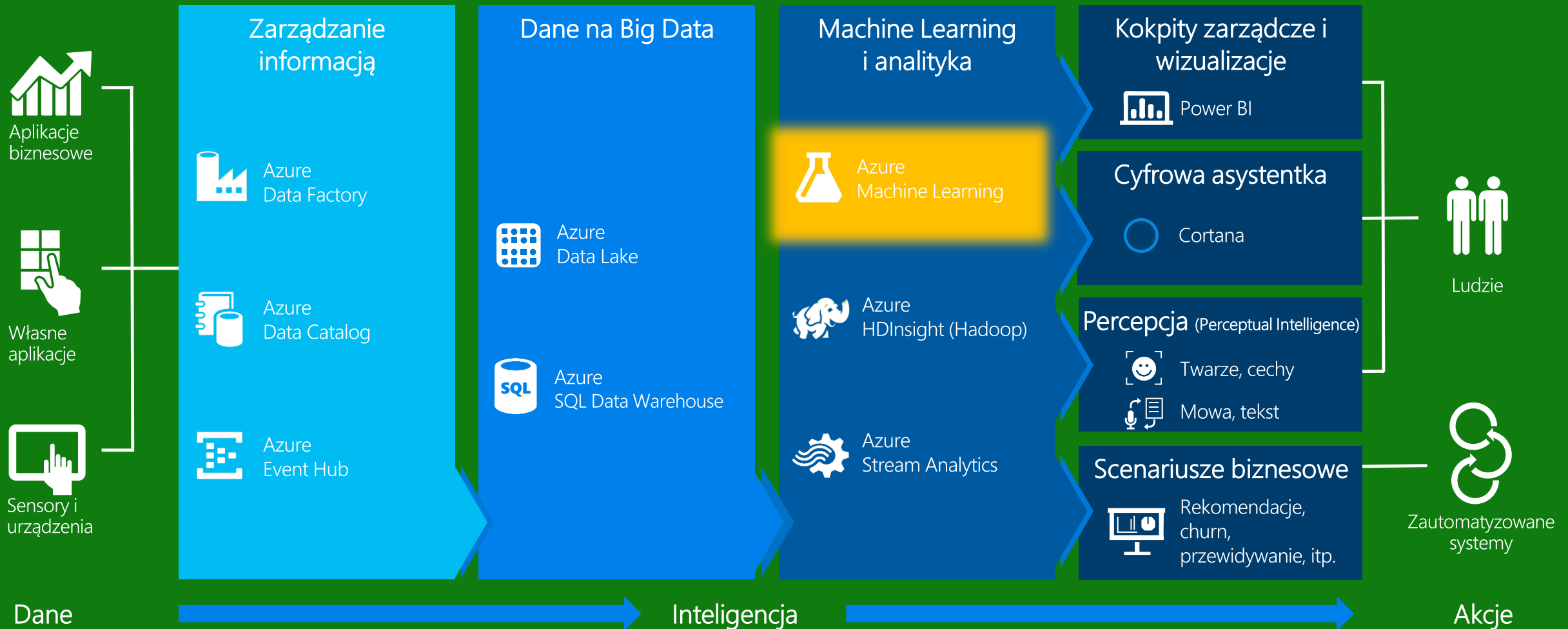
Rozwiązanie trudnych problemów

Większa wartość z Big Data

Zmiany w business analytics

# Cortana Analytics Suite

## Od danych do akcji



Dwa przykłady



getting a morgage in seattle

8,140,000 RESULTS Any time

Ads related to getting a morgage in seattle

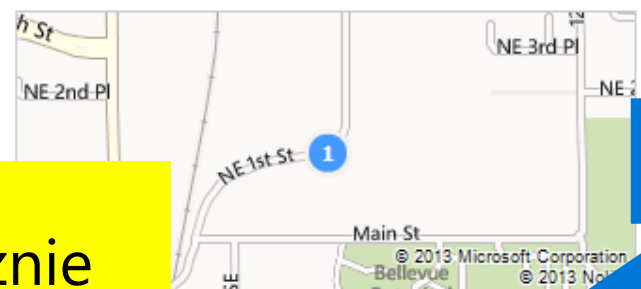
**15-Year Mortgage Rates | QuickenLoans.com**  
[www.QuickenLoans.com/Rates](http://www.QuickenLoans.com/Rates)  
 Lock Your Rate. 3.500% (3.92% APR) With America's #1 Online Lender.

**LendingTree®Official Site - Amazingly Low Mortgage Rates.**  
[LendingTree.com/Get-a-Mortgage](http://LendingTree.com/Get-a-Mortgage)  
 APR fr

**TILA**  
 seattle  
 Meet o

**Pre Q**  
 www.w  
 Estima

Seattle's Best Mortgage Inc



St Ste B306 - Bellevue  
0 - Directions  
lesbm.com

RELATED SEARCHES

- Getting a First Mortgage
- Getting a Mortgage Self-Employed
- Getting a Mortgage Loan Approved
- Getting a Mortgage On Land
- Getting a Mortgage in 2013
- How to Get a Mortgage License
- How to Get a Mortgage After Bankruptcy
- Mortgage Calculator

Ads related to getting a morgage in seattle

**Seattle Mortgage Rates**  
[Seattle.BankRateLocator.com](http://Seattle.BankRateLocator.com)  
 Rates Dropped to 3.18%. No Closing Costs! G  
 Quotes in 30 Seconds

12 Year Mortgage Rates

Nauczanie maszynowe praktycznie dotyka „wartości” każdego elementu wyniku wyszukiwania

Jaki to język

Jakie pokazać reklamy i w jakiej kolejności?

Źle napisane?

Które linki będą kliknięte?

Jakie jest prawdopodobieństwo kliknięcia danej reklamy

Co było celem?

Czy jakaś strona jest podejrzana

Które strony powinny być indeksowane?

Jaka cena za reklamę zapewni optymalny strumień dochodów?



Setki tysięcy maszyn...  
 Setki metryk i sygnałów z maszyny...  
 Co minutę/sekundę...

*Które sygnały są powiązane i wskazują realny problem?  
 Jak wiedzieć co naprawić WCZEŚNIEJ?*

# Azure Machine Learning

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow diagram for an experiment titled "02 TKFY16 R Anomaly Detection: Credit Risk". The workflow consists of several interconnected modules:

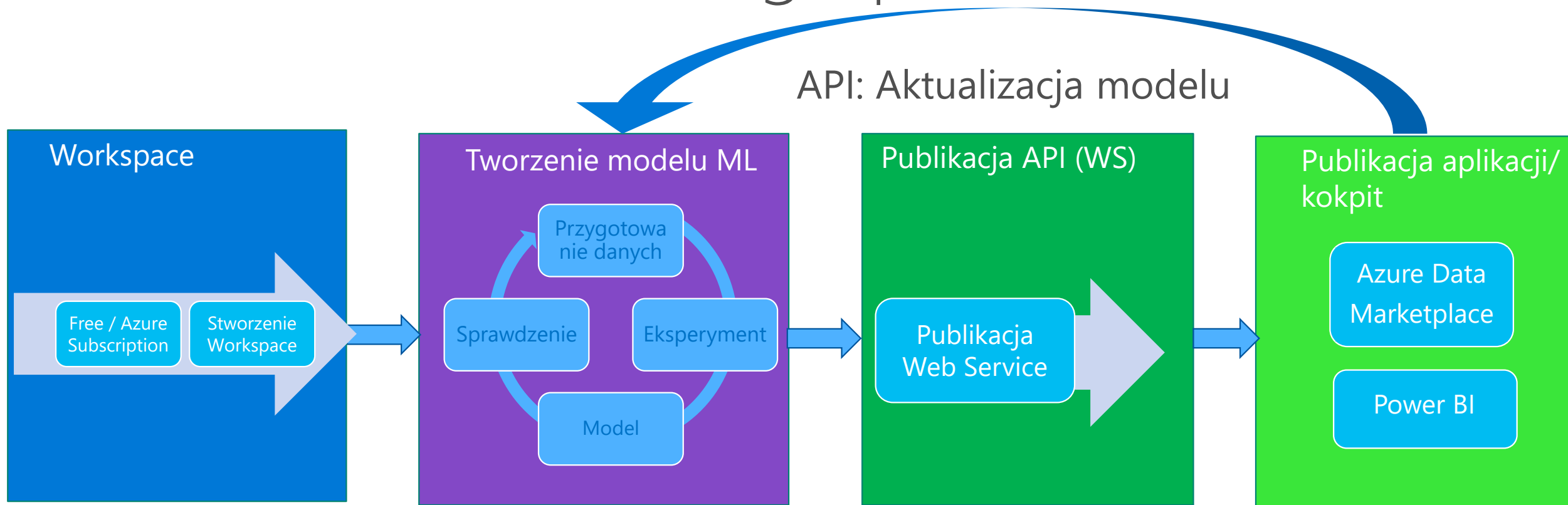
- Split**: Split into training (80%) and validation (20%) sets using stratified sampling.
- Split**: Select rows containing 'normal' transactions.
- One-Class Support Vector...**: Parameters:  $\nu = [1e-3, 1e-2, 1e-1, 1]$ ,  $\epsilon = [1e-4, 1e-3, 1e-2]$ .
- PCA-Based Anomaly Detecti...**: Parameters: rank = [2, 4, 6, 8, 10], oversampling = [2, 4, 6, 8, 10].
- Sweep Parameters**: optimize  $\nu$  and  $\epsilon$  parameters for maximum F-score.
- Sweep Parameters**: optimize rank and oversampling for maximum F-score.
- Train Anomaly Detection M...**: Train on 'normal' transactions only.
- Train Anomaly Dete**: Train on 'normal' tra only.

The right-hand side of the interface shows the **Properties** panel, which includes:

- Experiment Properties**: START TIME (9/23/2015...), END TIME (9/23/2015...), STATUS CODE (Finished), STATUS DETAILS (None).
- Summary**: Attempts to predict credit risk as anomalies within the data.
- Description**: Enter the detailed description for your experiment.
- [Original Experiment Documentation](#)
- Quick Help**

The bottom of the interface features a navigation bar with icons for NEW, RUN HISTORY, SAVE, DISCARD CHANGES, RUN, SET UP WEB SERVICE, and PUBLISH TO GALLERY.

# Azure Machine Learning - proces



<http://Azure.com/ml>

<https://Studio.azureml.net>  
<http://gallery.azureml.net>

<https://datamarket.azure.com/browse?query=machine+learning>

<https://powerbi.microsoft.com/>



# Jak ktoś chce zarobić,

Tu

home

home

Welcome, what would you like to publish?

- developer services 0
- virtual machines 0
- data services 0
- solution templates 0
- promotions 0
- publishers 0

Publish an  
DEVELOPER SERVICE



#### What is this?

Fully managed developer services (SAAS) for developers to use in custom application development.

#### What you provide

- Provisioning endpoint (API)
- Single sign-on endpoint
- Usage reporting endpoint

#### Licensing options

- *Fixed monthly subscription:* Billed based on plans defined by publisher (Azure bills on 30 day cycle based on subscription; customer must cancel plan to cease billing)
- *Consumption billing:* Billed based on how much your customer uses your service.

Microsoft bills Azure customer via Azure subscription on behalf of publisher.

Publish a  
VIRTUAL MACHINE



#### What is this?

Fully-configured virtual machine.

#### What you provide

- VM image (one or more VHD files) that is pre-configured to boot with fully installed application.

#### Licensing options

- *Consumption billing:* Azure billing automatically meters and customer will not be billed if they cease using the VM

Microsoft bills Azure customer via Azure subscription on behalf of publisher or purchased directly from publisher (BYOL).

Publish a  
DATA SERVICE



#### What is this?

Data services for use in custom application development by developers or by information workers in data analysis tools.

#### What you provide

- Database or Service endpoint
- Authentication information
- Service Mapping

#### Licensing options

- *Fixed monthly subscription:* Azure provides the customer with a limited number of transactions against your service per month.

Microsoft bills customer via Azure subscription or as separate credit card purchase on behalf of publisher.

Publish a  
SOLUTION TEMPLATE



#### What is this?

Solution templates allow you to combine multiple items already available in the marketplace to create more powerful solutions.

#### What you provide

- Azure Resource Manager (ARM) Template
- UI Definition file
- Optional installation scripts

#### Licensing options

Customers only pay for the individual components in the solution, based on how those are licensed.

# Przykłady:

MACHINE LEARNING API

## Speech APIs


Microsoft • published on April 22, 2015

### Description

Part of Microsoft Project Oxford, Speech APIs provide state-of-the-art algorithms to process spoken language. With these APIs developers can easily include the ability to add speech driven actions into their applications. In certain cases, the APIs also allow for real-time interaction with the user as well.


### Speech Recognition

Convert spoken audio to text. The API can be directed to turn on and recognize from the microphone in real-time, recognize from a different real-time audio source, or to recognize from audio in a file. In all cases, real-time streaming is available, so as the audio is being sent to the server, partial recognition results are returned.




### Text to Speech Conversion

Convert text to spoken audio. When applications need to *talk* back to their users, this API can be used to convert text generated by the app into audio that can be played back to the user.



Hashtag: #ProjectOxford



Sign up for free

Try it now

6675 views

[Tweet](#) [Share](#) [Email](#)

### LINKS

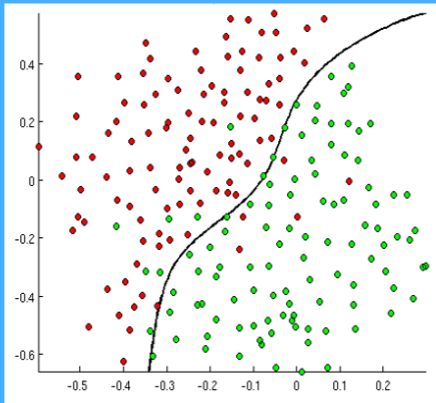
- [Project Oxford](#)
- [Publisher Offer Terms](#)
- [Publisher Offer Privacy Statement](#)

### TAGS

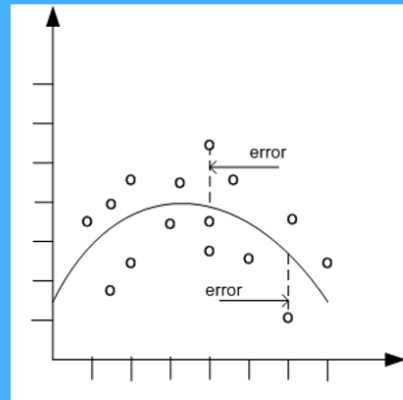
Speech Project Oxford Text Intent

# Klasy algorytmów w Azure ML

Classification



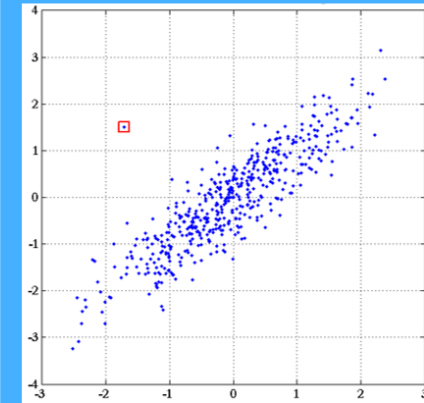
Regression



Recommenders



Anomaly Detection



Clustering



Dane numeryczne, tekstowe, obrazy, ....

# Który algorytm użyć?

To zależy (taa...)

[Cheat Scheet](#)

[How to choose algorithms for Microsoft Azure  
Machine Learning](#)

[BARDZO DOBRA DOKUMENTACJA](#)

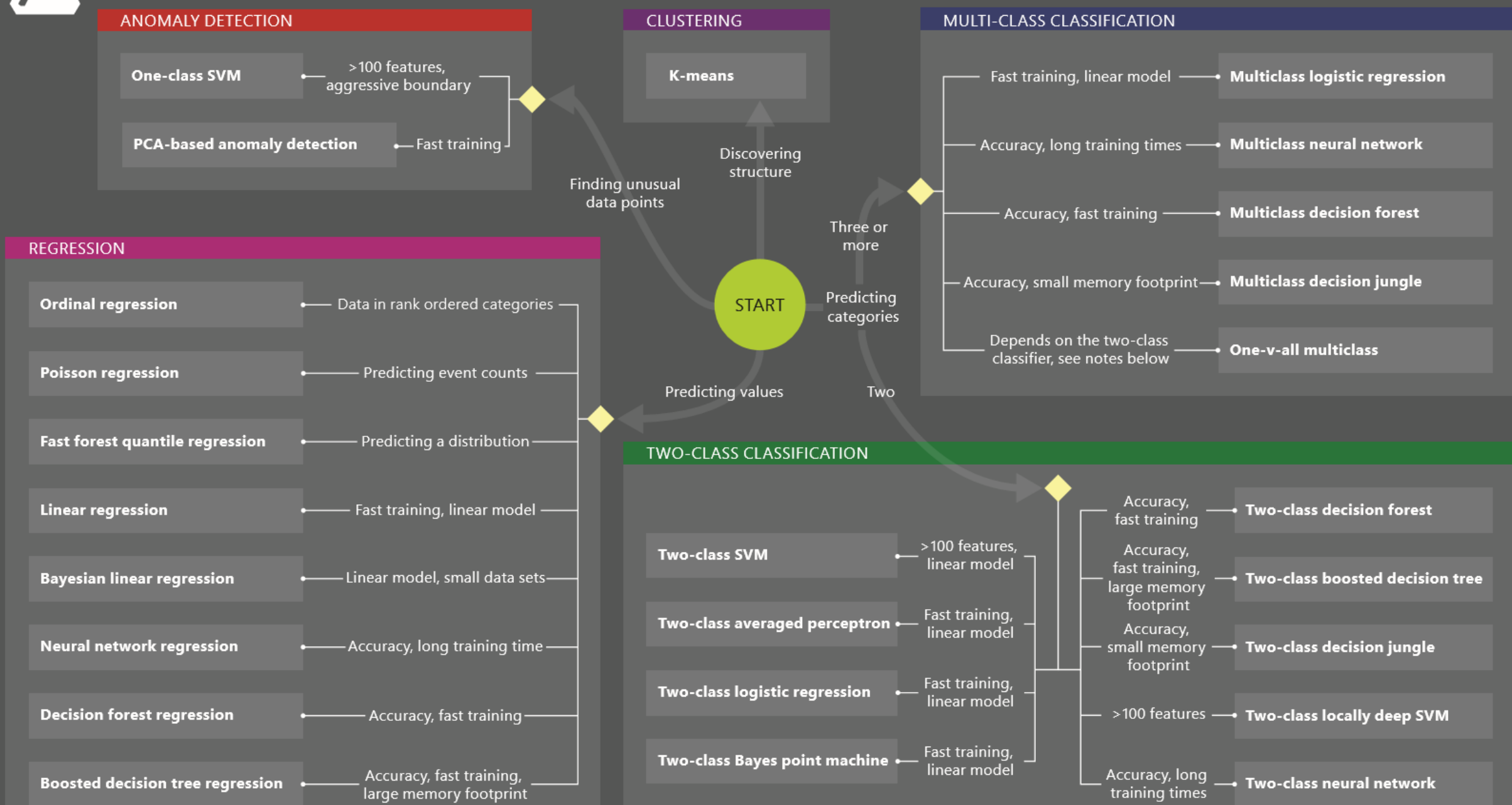
Instrukcja dla „developerów” jak użyć

Też odsyłająca do źródłowych prac naukowych



# Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.



Jest jeszcze Vowpal Wabbit

Vowpal Wabbit (John Langford) to szybki, równoległy framework machine learning rozwijany przy wsparciu Yahoo! Research a teraz Microsoft Research.

Cechy: (nie tłumaczę): continuous learning (online learning), dimensionality reduction, interactive learning.

Rozmiar problemu może przekraczać dostępną pamięć.

Rozwiązuje zarówno problemy klasyfikacji jak i regresji, z danymi częściowo etykietowanymi

# Inne możliwości ML

## Praca z danymi

Reader, writer, filtrowanie, uzupełnianie kolumn, konwersja, metadane

## Obsługa R (model, skrypty, R Device do wykresów)

484 pakiety, możliwość dodawania własnych modułów, ...

Możliwość serializacji wyjścia z R i wczytania „później”

Też skrypty dla miłośników Python

## Statystyki, funkcje matematyczne

## Analiza tekstu

Wybór cech/zależności (korelacja Pearsona, Chi-square, ...)

Zarządzanie historycznymi „uruchomieniami”

API! Integracja z Data Factory, HDInsight, ...

No to policzmy zdolność  
kredytową Niemców



# Jak szybko działa i ile kosztuje

[Cennik tu](#)

Istotne:

\$9.99 za Workspace (nie za użytkownika!)

\$1 za godzinę liczenia (trenowania) eksperymentu

API: \$2 / za godzinę liczenia + \$0.50 za 1000 wywołań

# Podsumowanie

Dużo gotowych algorytmów (które trzeba znać!)

Prosty do rozszerzania

Za darmo w ograniczonym zakresie

Nic nie trzeba „podawać” by zacząć!!!

Za darmo w subskrypcji Azure z MSDN / inne

Tani w produkcji (proporcjonalny do użycia)

Czyli – jak nam płacą za API to my płacimy coś dla Microsoft

Szansa zarobienia na nauce 😊