



Azure Machine Learning Studio

DAVID DAXBACHER

Agenda

- ▶ Aktuelles „Problem“
- ▶ Azure Machine Learning Studio
- ▶ Machine Learning – Algorithmen
- ▶ Erster Versuch – Klassifizierung
- ▶ Live – Demo (Recommender System)
- ▶ Retrain Model

„Problem“

bet-at-home.com

Acc-Nr./E-mail Password Forgot? Log in Register now! Guest login

SPORTS LIVE BETTING CASINO LIVE CASINO GAMES VIRTUAL POKER PROMOTIONS APPS

Search sports bets

Online betting news 2 / 5

Football
Tennis
Basketball
Handball
Volleyball
Ice Hockey
American Football
Motorsports
Politics
Baseball
Counter-Strike
Darts
League of Legends
Dota 2
Cycling
Winter sports
Rugby
Aussie Rules
Boxing
Combat Sports
StarCraft II
Floorball
Snooker
Table Tennis
Cricket
Golf
Society

NEWS!
CONTACT

Only one can win

The Green Bay Packers are looking to keep their home record clean with a win against the Chicago Bears. But also the guests are eager to win after celebrating the first victory of the season in their third game. Who will add a win at Lambeau Field?

Packers vs. Bears

Live betting Show all (830) 1 / 4

Zverev, Mischa - Sela, Dudi 0:1
LIVE 2nd set

Fritz, Taylor - Basilashvili, Nikoloz 1:0
LIVE 2nd set

Lu Y-H / Sharan D - Inglot D / Nestor D 1:0
LIVE 2nd set

Zvonareva, Vera - Krunic, Aleksandra 0:0
LIVE Interrupted

Fucsovics, Marton - Bublik, Alexander 0:0
LIVE 1st set

Sports betting highlights 1 X 2

Thiem, Dominic - Pella, Guido 1.27 -- 3.45
28.09.17 11:30

Zverev, Alexander - Darcis, Steve 1.14 -- 4.94
28.09.17 12:00

Muguruza, Garbiñe - Ostapenko, Jelena 1.39 -- 2.76
28.09.17 14:00

BATE Borisov - Arsenal FC 5.32 3.94 1.59
28.09.17 19:00

RB Salzburg - Olympique Marseille 2.35 3.37 2.88
28.09.17 19:00

Athletic Bilbao - FC Zorya Lugansk 1.26 5.40 11.16
28.09.17 19:00

Östersunds FK - Hertha BSC 2.97 3.18 2.39
28.09.17 19:00

OGC Nice - Vitesse Arnhem 1.44 4.42 6.79
28.09.17 19:00

Bet slip My bets

1-Click Single Combi System Multi...

Single bet

1 event - 1 tip! The single bet is the simplest bet offered by bet-at-home.com, with the fastest chance of winning.

To select a new bet, please click on the respective odds.

No betting account yet? Register now and get a bonus!

Open a betting account now!

Bonus 2 / 5

50 EUR SPORTS BONUS

You bet - we pay: get a bonus now!

Promotions 1 / 2

ON TARGET

Money back if no goals are scored!

Mini Blackjack

€1 Min 640 Max

INSURANCE PAYS 2 TO 1
BLACK JACK PAYS 3 TO 2
Dealer stands on 17 and must draw to 16

Football

Now live (2)

- #YourBet
- Champions League
- Europa League
- World Cup
- European Championship
- International Clubs
- Germany
- Germany Amateur
- England
- England Amateur
- Italy
- Italian amateurs
- Spain
- France
- Portugal
- Austria
- Poland
- Netherlands
- Belgium
- Scotland

Show all

- Tennis
- Basketball
- Handball

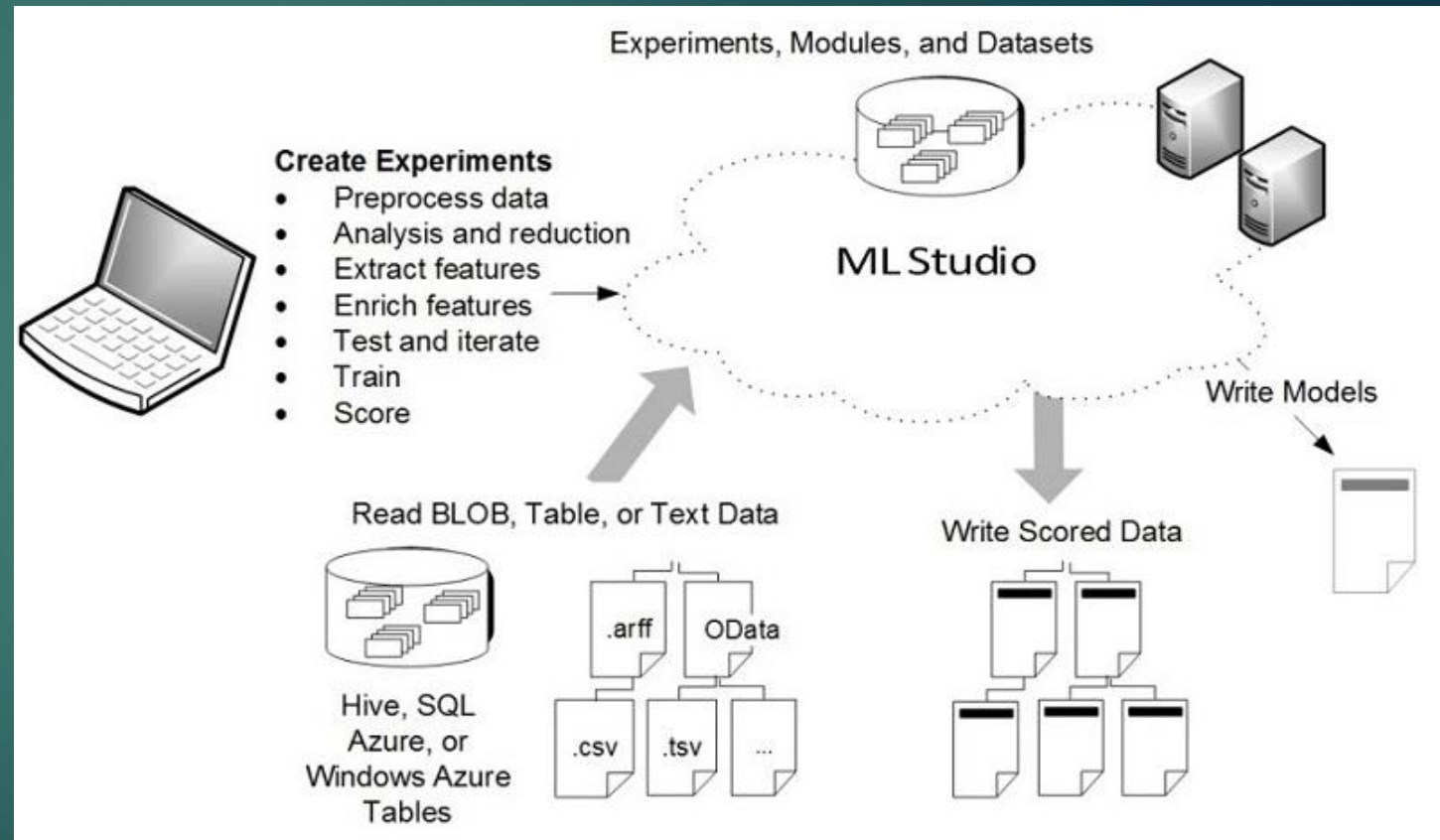
Football

Now live (2)

- #YourBet
- Champions League
- Europa League
- World Cup
- European Championship
- International Clubs
- Germany
- Germany Amateur
- England
 - Premier League
 - Premier League top match
 - Premier League Overall Winner
 - Premier League Relegated Team
 - Premier League Bottom Team
 - Premier League Top Goal scorer
 - Premier League Head-to-Head
 - Premier League Special
 - Chelsea FC, Special
 - Manchester United, Special
 - Manchester City, Special
 - Liverpool FC, Special
 - Championship
 - League One
 - League One Overall Winner
 - League Two
 - League Two Overall Winner
 - National League
 - National League, Overall Winner
 - FA Cup, Qualification
 - FA Cup Overall Winner
 - League Cup
 - League Cup Overall Winner
- England Amateur
- Italy
- Italian amateurs
- Spain
- France
- Portugal

Azure Machine Learning Studio

- ▶ Interaktives Drag and Drop Tool
- ▶ Fully managed
- ▶ Keine Programmierung erforderlich
- ▶ Trainingsexperiment → Vorhersageexperiment → Webservice



Machine Learning – Algorithmen

The screenshot displays the Microsoft Azure Machine Learning Studio interface. The main workspace shows a workflow diagram with three components: "Neural Network Regression", "Weather Dataset", and "Train Model". Arrows indicate that both the "Neural Network Regression" and "Weather Dataset" components are connected to the "Train Model" component. The left sidebar contains a search bar and a list of machine learning algorithms, with "Neural Network R..." and "Train Model" highlighted by red boxes. Red arrows point from these boxes to the corresponding components in the workflow diagram. The top right corner shows the user's name "DavidDaxbacherTest" and the status "In draft". The right sidebar contains the "Properties" panel, which includes sections for "Experiment Properties", "Summary", and "Description". The bottom of the interface features a toolbar with icons for "NEW", "RUN HISTORY", "SAVE", "SAVE AS", "DISCARD CHANGES", "RUN", "SET UP WEB SERVICE", and "PUBLISH TO GALLERY".

Microsoft Azure Machine Learning Studio

Experiment created on 10/2/2017

DavidDaxbacherTest

In draft

Draft saved at 1:03:30 PM

Properties Project

Experiment Properties

STATUS CODE InDraft

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

Enter the detailed description for your experiment.

Quick Help

Neural Network Regression

Weather Dataset

Train Model

Search experiment items

Feature Selection

Machine Learning

Evaluate

Initialize Model

Anomaly Detection

Classification

Clustering

Regression

Bayesian Linear R...

Boosted Decision ...

Decision Forest R...

Fast Forest Quanti...

Linear Regression

Neural Network R...

Ordinal Regression

Poisson Regression

Score

Train

Sweep Clustering

Train Anomaly Detect...

Train Clustering Model

Train Matchbox Reco...

Train Model

Tune Model Hyperpa...

OpenCV Library Modules

Python Language Modules

R Language Modules

NEW

RUN HISTORY

SAVE

SAVE AS

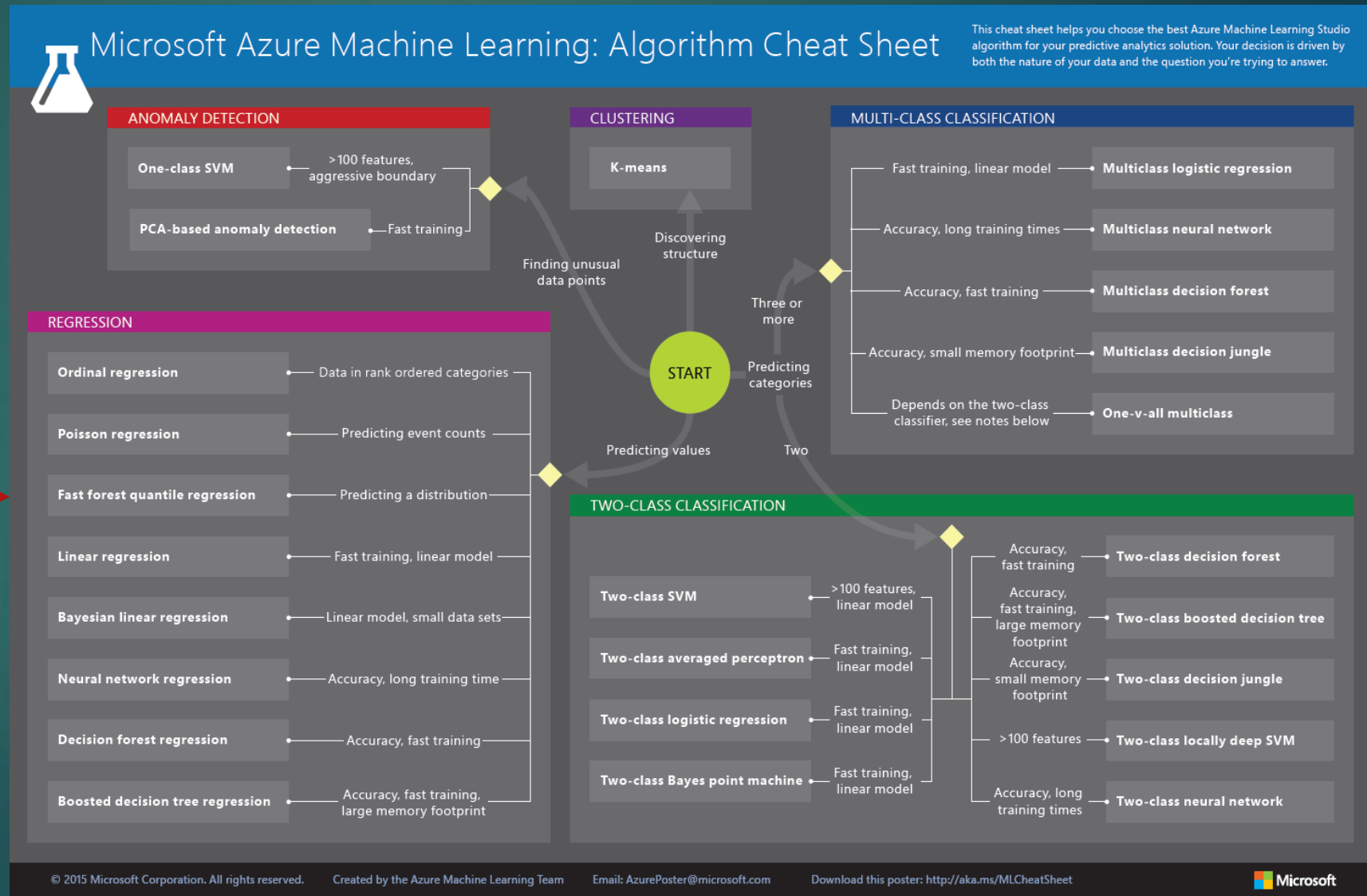
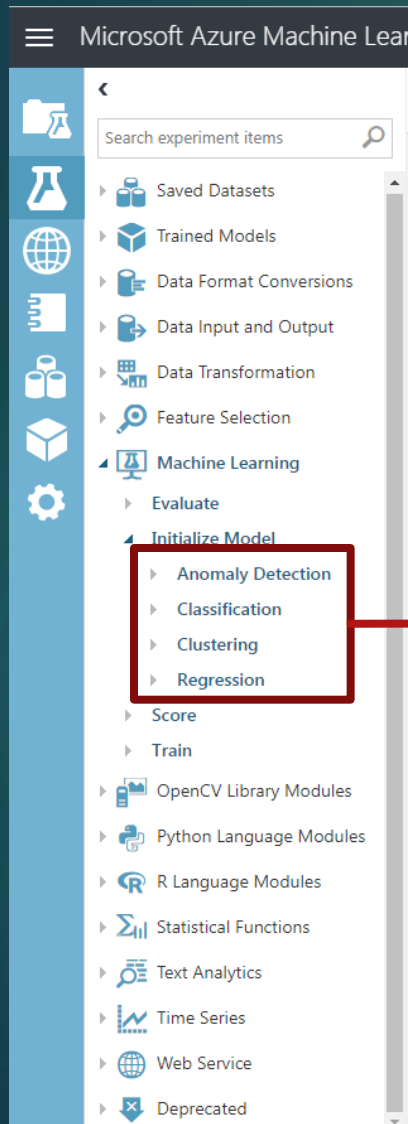
DISCARD CHANGES

RUN

SET UP WEB SERVICE

PUBLISH TO GALLERY

Machine Learning – Algorithmen



Machine Learning – Algorithmen

Microsoft Azure Machine Learning

Search experiment items

- Saved Datasets
- Trained Models
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
 - Evaluate
 - Initialize Model
 - Anomaly Detection**
 - Classification
 - Clustering
 - Regression
 - Score
 - Train
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series
- Web Service
- Deprecated

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.

ANOMALY DETECTION

- One-class SVM — > 100 features, aggressive boundary
- PCA-based anomaly detection — Fast training

CLUSTERING

- K-means

MULTI-CLASS CLASSIFICATION

- Fast training, linear model — Multiclass logistic regression
- Accuracy, long training times — Multiclass neural network
- Accuracy, fast training — Multiclass decision forest
- Accuracy, small memory footprint — Multiclass decision jungle
- Depends on the two-class classifier, see notes below — One-v-all multiclass

REGRESSION

- Ordinal regression
- Poisson regression
- Fast forest quantile
- Linear regression
- Bayesian linear regression
- Neural network regression — Accuracy, long training time
- Decision forest regression — Accuracy, fast training
- Boosted decision tree regression — Accuracy, fast training, large memory footprint

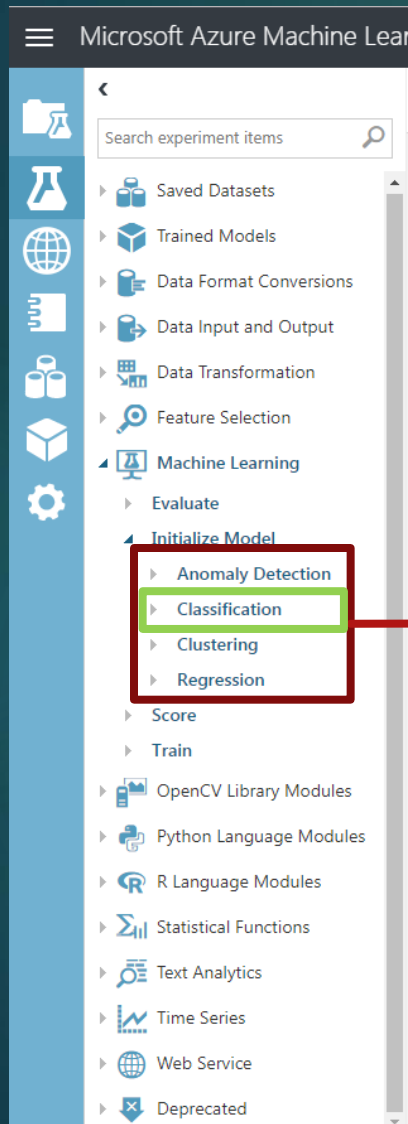
TWO-CLASS CLASSIFICATION

- Two-class SVM — > 100 features, linear model
- Two-class averaged perceptron — Fast training, linear model
- Two-class logistic regression — Fast training, linear model
- Two-class Bayes point machine — Fast training, linear model
- Accuracy, fast training — Two-class decision forest
- Accuracy, fast training, large memory footprint — Two-class boosted decision tree
- Accuracy, small memory footprint — Two-class decision jungle
- > 100 features — Two-class locally deep SVM
- Accuracy, long training times — Two-class neural network

Anomaly Detection: Credit Risk

© 2015 Microsoft Corporation. All rights reserved. Created by the Azure Machine Learning Team Email: AzurePoster@microsoft.com Download this poster: <http://aka.ms/MLCheatSheet> Microsoft

Machine Learning – Algorithmen



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.

ANOMALY DETECTION

- One-class SVM: > 100 features, aggressive boundary
- PCA-based anomaly detection: Fast training

REGRESSION

- Ordinal regression: Data in rank ordered categories
- Poisson regression: Predicting event counts
- Fast forest quantile regression: Predicting a distribution
- Linear regression
- Bayesian linear regression
- Neural network regression
- Decision forest regression
- Boosted decision tree regression

Compare Multi-class Classifiers: Letter recognition

Letter	Probability
a	0.7
q	0.2
o	0.1

MULTI-CLASS CLASSIFICATION

- Fast training, linear model: Multiclass logistic regression
- Accuracy, long training times: Multiclass neural network
- Accuracy, fast training: Multiclass decision forest
- Accuracy, small memory footprint: Multiclass decision jungle
- Depends on the two-class classifier, see notes below: One-v-all multiclass

Binary Classification: Breast cancer detection

TWO-CLASS CLASSIFICATION

- Two-class SVM: > 100 features, linear model
- Two-class averaged perceptron: Fast training, linear model
- Two-class logistic regression: Fast training, linear model
- Two-class Bayes point machine: Fast training, linear model
- Two-class decision forest: Accuracy, fast training
- Two-class boosted decision tree: Accuracy, fast training, large memory footprint
- Two-class decision jungle: Accuracy, small memory footprint
- Two-class locally deep SVM: > 100 features
- Two-class neural network: Accuracy, long training times

© 2015 Microsoft Corporation. All rights reserved. Created by the Azure Machine Learning Team. Email: AzurePoster@microsoft.com. Download this poster: <http://aka.ms/MLCheatSheet>

Machine Learning – Algorithmmen

Microsoft Azure Machine Learning

Search experiment items

- Saved Datasets
- Trained Models
- Data Format Conversions
- Data Input and Output
- Data Transformation
- Feature Selection
- Machine Learning
 - Evaluate
 - Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
 - Score
 - Train
- OpenCV Library Modules
- Python Language Modules
- R Language Modules
- Statistical Functions
- Text Analytics
- Time Series
- Web Service
- Deprecated

Clustering: Color quantization

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.

START

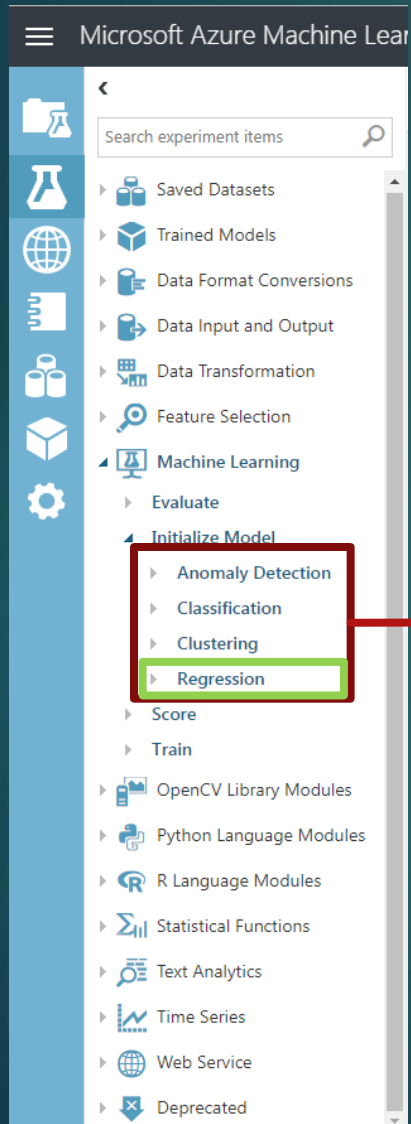
- CLUSTERING**
 - Discovering structure
 - Identifying unusual data points
 - Algorithm: **K-means**
- MULTI-CLASS CLASSIFICATION**
 - Fast training, linear model → **Multiclass logistic regression**
 - Accuracy, long training times → **Multiclass neural network**
 - Accuracy, fast training → **Multiclass decision forest**
 - Accuracy, small memory footprint → **Multiclass decision jungle**
 - Depends on the two-class classifier, see notes below → **One-v-all multiclass**
- TWO-CLASS CLASSIFICATION**
 - Predicting values
 - Predicting categories
 - Two
 - Three or more
 - Accuracy, fast training → **Two-class decision forest**
 - Accuracy, fast training, large memory footprint → **Two-class boosted decision tree**
 - Accuracy, small memory footprint → **Two-class decision jungle**
 - >100 features → **Two-class locally deep SVM**
 - Accuracy, long training times → **Two-class neural network**
 - >100 features, linear model → **Two-class SVM**
 - Fast training, linear model → **Two-class averaged perceptron**
 - Fast training, linear model → **Two-class logistic regression**
 - Fast training, linear model → **Two-class Bayes point machine**

REGRESSION

- Ordinal regression** — Data in rank ordered categories
- Poisson regression** — Predicting event counts
- Fast forest quantile regression** — Predicting a distribution
- Linear regression** — Fast training, linear model
- Bayesian linear regression** — Linear model, small data sets
- Neural network regression** — Accuracy, long training time
- Decision forest regression** — Accuracy, fast training
- Boosted decision tree regression** — Accuracy, fast training, large memory footprint

© 2015 Microsoft Corporation. All rights reserved. Created by the Azure Machine Learning Team Email: AzurePoster@microsoft.com Download this poster: <http://aka.ms/MLCheatSheet> Microsoft

Machine Learning – Algorithmen



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.

ANOMALY DETECTION

- One-class SVM → > 100 features, aggressive boundary
- PCA-based anomaly detection → Fast training

CLUSTERING

- K-means → Discovering structure

MULTI-CLASS CLASSIFICATION

- Fast training, linear model → Multiclass logistic regression
- Accuracy, long training times → Multiclass neural network
- Accuracy, fast training → Multiclass decision forest
- Accuracy, small memory footprint → Multiclass decision jungle
- Depends on the two-class classifier, see notes below → One-v-all multiclass

REGRESSION

- Ordinal regression → Data in rank ordered categories
- Poisson regression → Predicting event counts
- Fast forest quantile regression → Predicting a distribution
- Linear regression → Fast training, linear model
- Bayesian linear regression → Linear model, small data sets
- Neural network regression → Accuracy, long training time
- Decision forest regression → Accuracy, fast training
- Boosted decision tree regression → Accuracy, fast training, large memory footprint

Regression: Demand estimation

TWO-CLASS CLASSIFICATION

- Accuracy, fast training → Two-class decision forest
- Accuracy, fast training, large memory footprint → Two-class boosted decision tree
- Accuracy, small memory footprint → Two-class decision jungle
- > 100 features → Two-class locally deep SVM
- Accuracy, long training times → Two-class neural network

Other Regression Algorithms

- Two-class logistic regression → Fast training, linear model
- Two-class Bayes point machine → Fast training, linear model

© 2015 Microsoft Corporation. All rights reserved. Created by the Azure Machine Learning Team Email: AzurePoster@microsoft.com Download this poster: <http://aka.ms/MLCheatSheet> Microsoft

Classification

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9	Col10	Col11	Col12	Col13	Col14
T	2	8	3	5	1	8	13	0	6	6	10	8	0	
I	5	12	3	7	2	10	5	5	4	13	3	9	2	
D	4	11	6	8	6	10	6	2	6	10	3	7	3	
N	7	11	6	6	3	5	9	4	6	4	4	10	6	
G	2	1	3	1	1	8	6	6	6	6	5	9	1	
S	4	11	5	8	3	8	8	6	9	5	6	6	0	
B	4	2	5	4	4	8	7	6	6	7	6	6	2	
A	1	1	3	2	1	8	2	2	2	8	2	8	1	
J	2	2	4	4	2	10	6	2	6	12	4	8	1	
M	11	15	13	9	7	13	2	6	2	12	1	9	8	
X	3	9	5	7	4	8	7	3	8	5	6	8	2	
O	6	13	4	7	4	6	7	6	3	10	7	9	5	
G	4	9	6	7	6	7	8	6	2	6	5	11	4	
M	6	9	8	6	9	7	8	6	5	7	5	8	8	
R	5	9	5	7	6	6	11	7	3	7	3	9	2	
F	6	9	5	4	3	10	6	3	5	10	5	7	3	
O	3	4	4	3	2	8	7	7	5	7	6	8	2	
C	7	10	5	5	2	6	8	6	8	11	7	11	2	
T	6	11	6	8	5	6	11	5	6	11	9	4	3	
J	2	2	3	3	1	10	6	3	6	12	4	9	0	
J	1	3	2	2	1	8	8	2	5	14	5	8	0	
H	4	5	5	4	4	7	7	6	6	7	6	8	3	

J

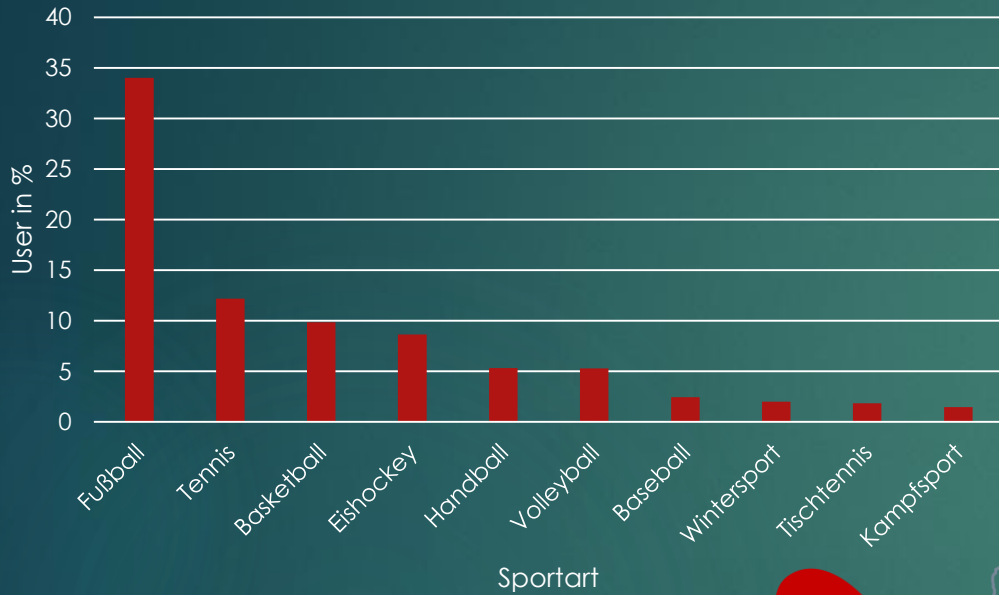
J

J

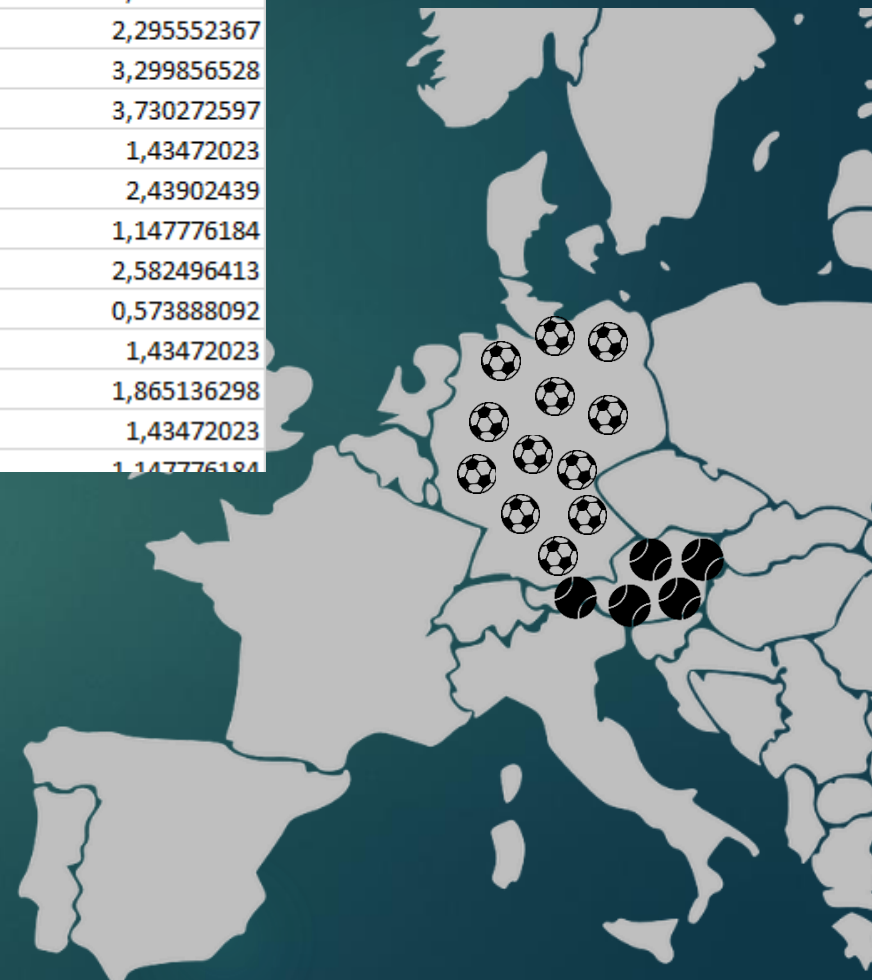
	SportID	Sex	CountryID	Street	ZIPCode	Birthdate
	Fußball	0	152	Street12108530	2632	1977-10-01T00:00:00
	Eishockey	0	75	Street12108541	51467	1945-05-04T00:00:00
	Fußball	0	75	Street12108541	51467	1945-05-04T00:00:00
	Tennis	0	75	Street12108541	51467	1945-05-04T00:00:00
	Fußball	0	75	Street12108563	63571	1975-09-30T00:00:00
	Streethockey	0	75	Street12108594	55262	1944-01-07T00:00:00
	Curling	0	75	Street12108594	55262	1944-01-07T00:00:00
	Wasserball	0	75	Street12108594	55262	1944-01-07T00:00:00
	Wintersport	0	75	Street12108594	55262	1944-01-07T00:00:00
	Feldhockey	0	75	Street12108594	55262	1944-01-07T00:00:00
	Tischtennis	0	75	Street12108594	55262	1944-01-07T00:00:00
	Floorball	0	75	Street12108594	55262	1944-01-07T00:00:00
	Pesäpallo	0	75	Street12108594	55262	1944-01-07T00:00:00
	Beachvolleyball	0	75	Street12108594	55262	1944-01-07T00:00:00
	Rugby	0	75	Street12108594	55262	1944-01-07T00:00:00
	Basketball	0	75	Street12108594	55262	1944-01-07T00:00:00
	Eishockey	0	75	Street12108594	55262	1944-01-07T00:00:00
	Bandy	0	75	Street12108594	55262	1944-01-07T00:00:00

Classification

User pro Sportart



Sportart	Durchschnittl. Wetten in Prozent
Fußball	28,69440459
Tennis	9,899569584
Basketball	4,447632712
Eishockey	5,595408895
Handball	2,295552367
Volleyball	3,299856528
Baseball	3,730272597
Wintersport	1,43472023
Tischtennis	2,43902439
Kampfsport	1,147776184
E-Sport	2,582496413
Boxen	0,573888092
Darts	1,43472023
Counter-Strike	1,865136298
Beachvolleyball	1,43472023
Rugby	1,147776184



Experiment created on 10/2/2017

In draft

Draft saved at 1:03:30 PM

Properties Project

Experiment Properties

STATUS CODE InDraft

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Description

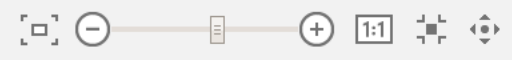
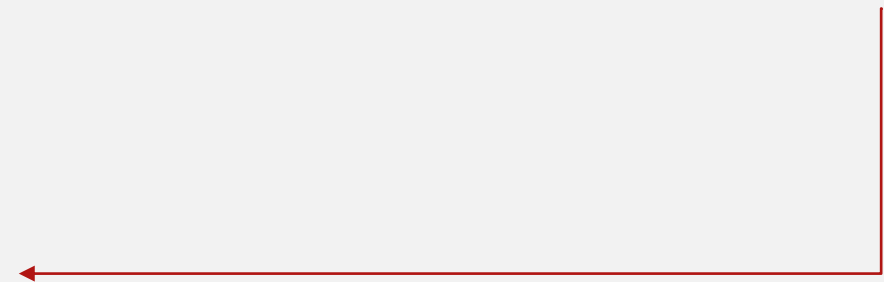
Enter the detailed description for your experiment.

Quick Help

- Search experiment items
- Feature Selection
- Machine Learning
 - Evaluate
 - Initialize Model
 - Anomaly Detection
 - Classification
 - Clustering
 - Regression
 - Bayesian Linear R...
 - Boosted Decision ...
 - Decision Forest R...
 - Fast Forest Quanti...
 - Linear Regression
 - Neural Network R...
 - Ordinal Regression
 - Poisson Regression
 - Score
 - Train
 - Sweep Clustering
 - Train Anomaly Detect...
 - Train Clustering Model
 - Train Matchbox Reco...
 - Train Model
 - Tune Model Hyperpa...
 - OpenCV Library Modules
 - Python Language Modules
 - R Language Modules

-- LIVE DEMO --

Recommender System



Evaluation - Rating Prediction

...
12345	Bundesliga	10
12345	NBA	3
12345	Serie A	5
...

12345	Bundesliga	10
12345	NBA	3

Trainingsdaten

12345	Serie A	5
-------	---------	---

Testdaten

Train Matchbox Recommender

12345	Serie A	???
-------	---------	-----

Evaluation - Rating Prediction

...
12345	Bundesliga	10
12345	NBA	3
12345	Serie A	5
...

12345	Bundesliga	10
12345	NBA	3

Trainingsdaten

12345	Serie A	5
-------	---------	---

Testdaten

Train Matchbox Recommender

12345	Serie A	6
-------	---------	---

Evaluation – Item Recommendation

...
12345	Bundesliga	10
12345	NBA	3
12345	Serie A	5
12345	Premier League	9
...

12345	Bundesliga	10
12345	Serie A	5

Trainingsdaten

12345	NBA	3
12345	Premier League	9

Testdaten

Train Matchbox Recommender

???	???			
-----	-----	--	--	--

Evaluation – Item Recommendation

...
12345	Bundesliga	10
12345	NBA	3
12345	Serie A	5
12345	Premier League	9
...

12345	Bundesliga	10
12345	Serie A	5

Trainingsdaten

12345	NBA	3
12345	Premier League	9

Testdaten

Train Matchbox Recommender

Premier League	NBA			
----------------	-----	--	--	--

Retrain Model

