

31.8.–3.9.2015
in Nürnberg



Herbstcampus

Wissenstransfer
par excellence

Bring your own Schufa

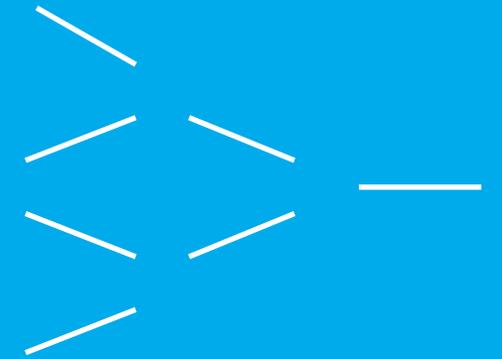
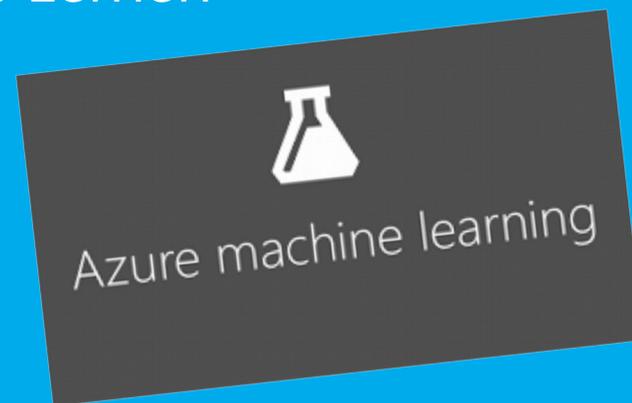
Prädikative Analyse selbst gemacht

Jan Schweda

complement AG

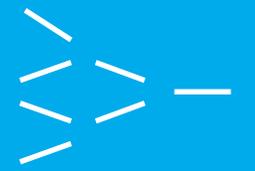
BRING YOUR OWN SCHUFAS!

Prädiktive Analysen durch Maschinelles Lernen



- ▶ Jan Schweda
- ▶ jan.schweda@complement.de
- ▶ Senior Softwareengineer Web & Cloud

WER BIN ICH?



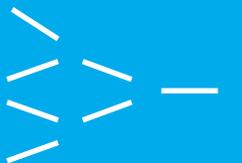


- ▶ Die Möglichkeiten von maschinellem Lernen aufzeigen
- ▶ Azure Machine Learning als Dienst vorstellen
- ▶ Die Schritte zum Erstellen und Veröffentlichen von Experimenten zeigen



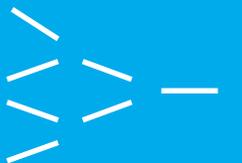
- ▶ Data Science oder BI Experten erschaffen
- ▶ Diskussionen über Datenschutz und Ethik führen

ZIELE DES VORTRAGS

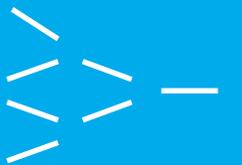


- ▶ Was ist Machine Learning?
- ▶ Was brauche ich um mit ML zu beginnen?
- ▶ Ein Experiment durchführen
- ▶ Und jetzt?

AGENDA



WAS IST MACHINE LEARNING?



Formale Definition

„A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E “

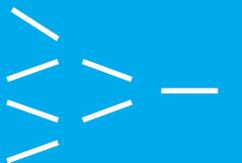
Tom M. Mitchell, Machine Learning, McGraw Hill

Lesbare Definition

„The goal of machine learning is to program computers to use example data or past experience to solve a given problem.“

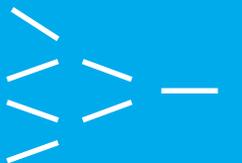
Introduction to Machine Learning, 3rd Edition, MIT Press

DEFINITION MACHINE LEARNING



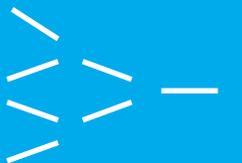
- ▶ Prädiktive Analysen
 - ▶ Überwachtes Lernen (supervised learning)
 - ▶ Erlernen einer Funktion aus gegebenen Daten (Eingaben und Ausgaben)
- ▶ Klassifizierung und Gruppierung
 - ▶ Nicht Überwachtes Lernen (unsupervised learning)
 - ▶ Clustering-Verfahren
 - ▶ Die Daten werden in mehrere Kategorien eingeteilt
 - ▶ Kategorien werden selbständig ermittelt

WOFÜR BRAUCHE ICH ML



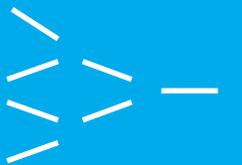
- ▶ Azure ML
 - ▶ Verwalteter Cloud Service
 - ▶ Keine zusätzliche Software notwendig
 - ▶ Keine eigene Hardware notwendig
- ▶ ML Studio
 - ▶ IDE für Machine Learning
 - ▶ Experimente als Datenflussdiagramme
 - ▶ Drag und Drop Editor für Experimente
 - ▶ R und Python Unterstützung

WAS IST AZURE ML?

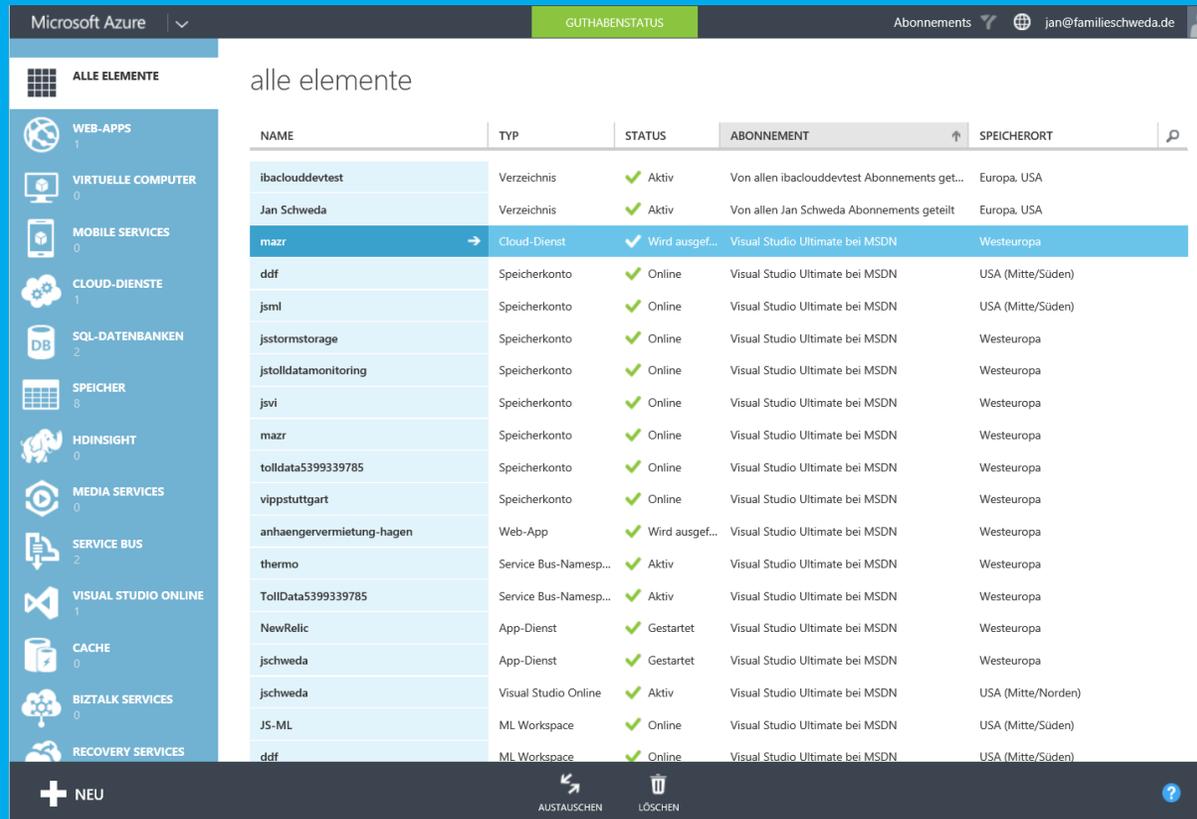


WAS BRAUCHE ICH UM MIT ML ZU BEGINNEN?

Azure Grundlagen



► manage.windowsazure.net

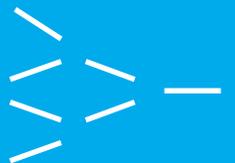


The screenshot displays the Microsoft Azure portal interface. The top navigation bar includes 'Microsoft Azure', a dropdown arrow, 'GUTHABENSTATUS', 'Abonnements', a globe icon, and the user email 'jan@familieschweda.de'. The left sidebar lists various service categories with their respective counts: WEB-APPS (1), VIRTUELLE COMPUTER (0), MOBILE SERVICES (0), CLOUD-DIENSTE (1), SQL-DATENBANKEN (2), SPEICHER (8), HDINSIGHT (0), MEDIA SERVICES (0), SERVICE BUS (2), VISUAL STUDIO ONLINE (1), CACHE (0), BIZTALK SERVICES (0), and RECOVERY SERVICES (0). The main content area, titled 'alle elemente', shows a table of resources. The 'mazr' resource is highlighted in blue. The table columns are NAME, TYP, STATUS, ABONNEMENT, and SPEICHERORT.

NAME	TYP	STATUS	ABONNEMENT	SPEICHERORT
ibacclouddevtest	Verzeichnis	✓ Aktiv	Von allen ibacclouddevtest Abonnements get...	Europa, USA
Jan Schweda	Verzeichnis	✓ Aktiv	Von allen Jan Schweda Abonnements geteilt	Europa, USA
mazr	Cloud-Dienst	✓ Wird ausgef...	Visual Studio Ultimate bei MSDN	Westeuropa
ddf	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	USA (Mitte/Süden)
jsml	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	USA (Mitte/Süden)
jsstormstorage	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	Westeuropa
jsolladatamonitoring	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	Westeuropa
jsvi	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	Westeuropa
mazr	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	Westeuropa
tolldata5399339785	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	Westeuropa
vippstuttgart	Speicherkonto	✓ Online	Visual Studio Ultimate bei MSDN	Westeuropa
anhaengervermietung-hagen	Web-App	✓ Wird ausgef...	Visual Studio Ultimate bei MSDN	Westeuropa
thermo	Service Bus-Namesp...	✓ Aktiv	Visual Studio Ultimate bei MSDN	Westeuropa
TollData5399339785	Service Bus-Namesp...	✓ Aktiv	Visual Studio Ultimate bei MSDN	Westeuropa
NewRelic	App-Dienst	✓ Gestartet	Visual Studio Ultimate bei MSDN	Westeuropa
jschweda	App-Dienst	✓ Gestartet	Visual Studio Ultimate bei MSDN	Westeuropa
jschweda	Visual Studio Online	✓ Aktiv	Visual Studio Ultimate bei MSDN	USA (Mitte/Norden)
JS-ML	ML Workspace	✓ Online	Visual Studio Ultimate bei MSDN	USA (Mitte/Süden)
ddf	ML Workspace	✓ Online	Visual Studio Ultimate bei MSDN	USA (Mitte/Süden)

At the bottom of the interface, there are icons for '+ NEU', 'AUSTAUSCHEN', 'LÖSCHEN', and a help icon.

AZURE PORTAL



NEU

	SQL-DATENBANK		
	SPEICHER		
	HDINSIGHT		
	RECOVERY SERVICES		
	MACHINE LEARNING		
	STREAM ANALYTICS		
	STORSIMPLE MANAGER NICHT VERFÜGBAR		

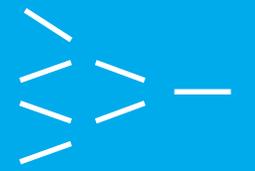
URL

STANDORT/AFFINITÄTSGRUPPE
USA (Mitte/Süden) ▼

REPLIKATION
Georedundant ▼

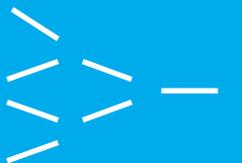
SPEICHERKONTO ERSTELLEN ✓

AZURE STORAGE ACCOUNT

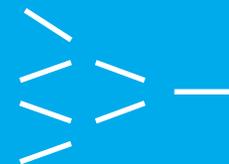


- ▶ <http://azure.microsoft.com/de-de/pricing/details/machine-learning/>
- ▶ <http://azure.microsoft.com/de-de/pricing/details/storage/>

AZURE PREISE

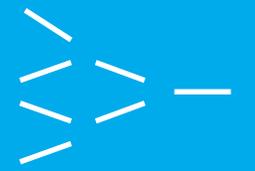


DAS EXPERIMENT



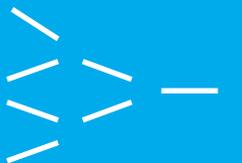
- ▶ Ein neues oder vorgegebenes Dataset hinzufügen
- ▶ Ein neues Experiment starten
- ▶ Die Daten per DnD hinzufügen
- ▶ Daten im ML Studio visualisieren

DATEN IMPORTIEREN



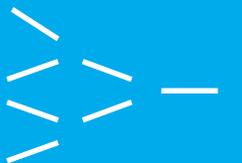
- ▶ Daten müssen aufbereitet werden bevor sie analysiert werden
 - ▶ Auf fehlende Werte muss reagiert werden
 - ▶ Nicht relevante Spalten sollen das Ergebnis nicht beeinflussen
- ▶ Daten müssen auf mindestens zwei Töpfe verteilt werden

DATEN AUFBEREITEN



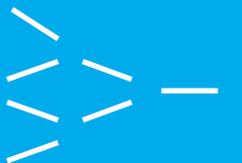
- ▶ Jede Zeile der Daten stellt eine Person dar
- ▶ Jede Spalte eine Eigenschaft
- ▶ Das Model wird aus den Eigenschaften der Daten erstellt
- ▶ Ein gutes Model braucht Erfahrung und Experimente

FÄHIGKEITEN DES MODELLS DEFINIEREN

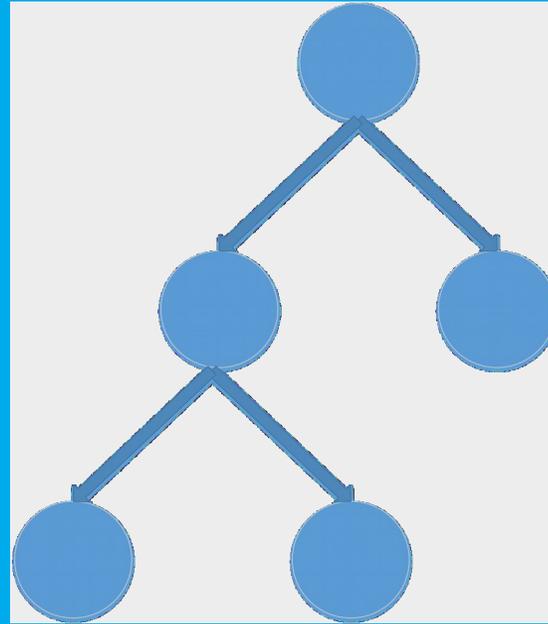


- ▶ Das Model muss trainiert und validiert werden
- ▶ Hierzu werden die Daten auf zwei Töpfe verteilt
- ▶ Über Machine Learning können Module zum initialisieren des Algorithmus ausgewählt werden

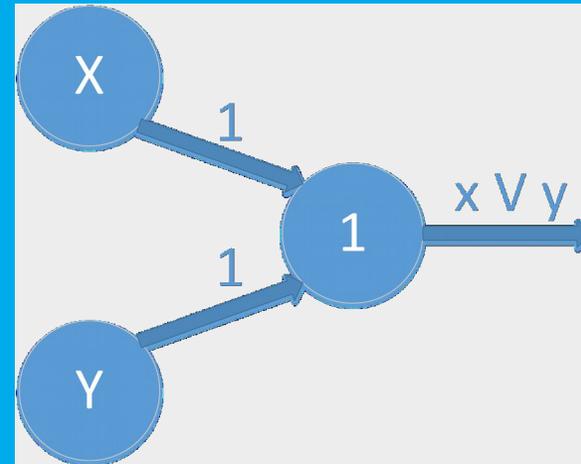
EINEN ALGORITHMUS AUSWÄHLEN



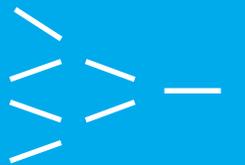
▶ Entscheidungsbäume



▶ Künstliche neuronale Netze

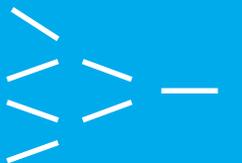


ALGORITHMEN

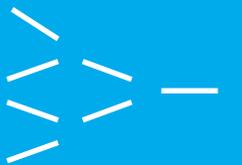


- ▶ Das trainierte Model muss mit Daten bewertet werden
- ▶ Die Ergebnisse der Bewertung:
 - ▶ Mean Absolute Error (MAE)
 - ▶ Root Mean Squared Error (RMSE)
 - ▶ Relative Absolute Error
 - ▶ Relative Squared Error
 - ▶ Coefficient of Determination
- ▶ Für alle Werten gilt, desto kleiner umso besser
- ▶ Bei „Coefficient of Determination“ ist 1.0 der Beste Wert

DAS MODELL TRAINIEREN UND AUSWERTEN

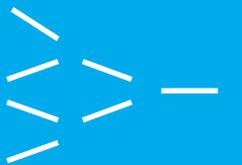


UND JETZT?





DAS EXPERIMENT ALS DIENST VERÖFFENTLICHEN



► Danke fürs zuhören!

