

Recommendation systems with Azure ML and Cognitive Services



Alexander Klein

Who am I ?

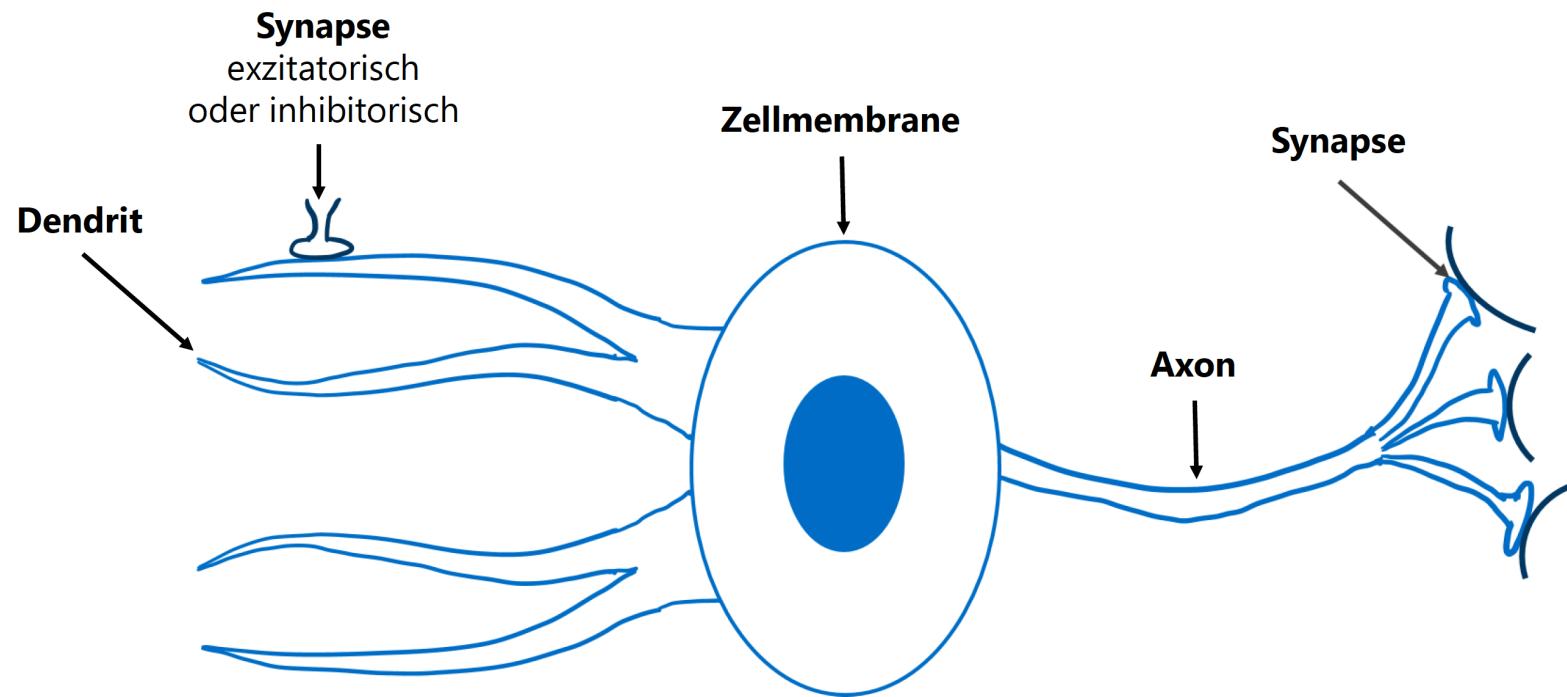
- Independent BI Consultant
 - > 15 years experience of SQL Server
 - Focus on BI Stack
 - BI on Azure
 - a.klein@consulting-bi.de
 - @SQL_Alex
 - consulting-bi.de

Next 60 minutes

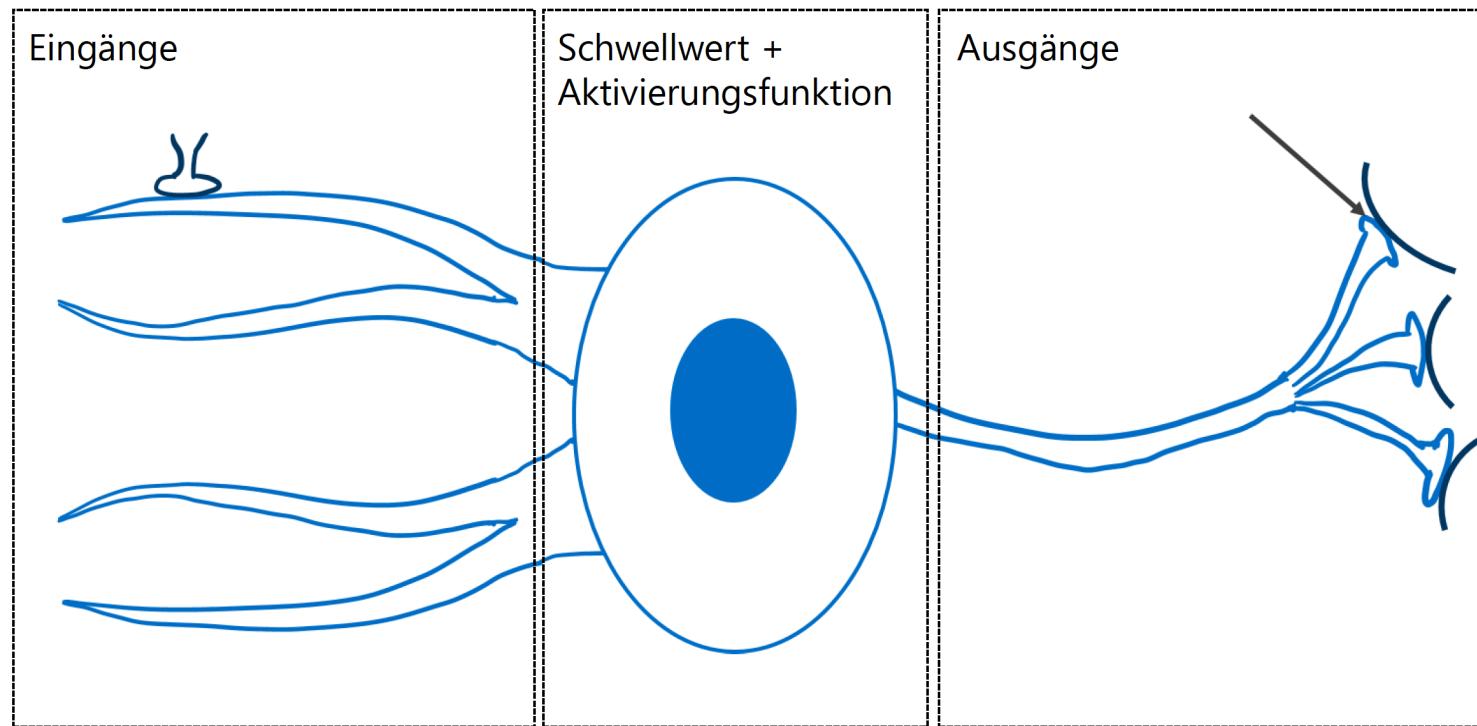
- Cognitive Services
- What war recommendation systems?
- Where are they used?
- Demo Cognitive Services
- Demo Azure ML
- Review
- Next steps



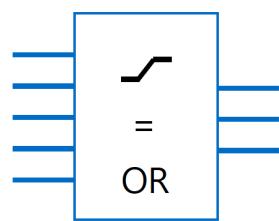
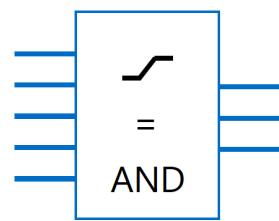
Cognitive Services



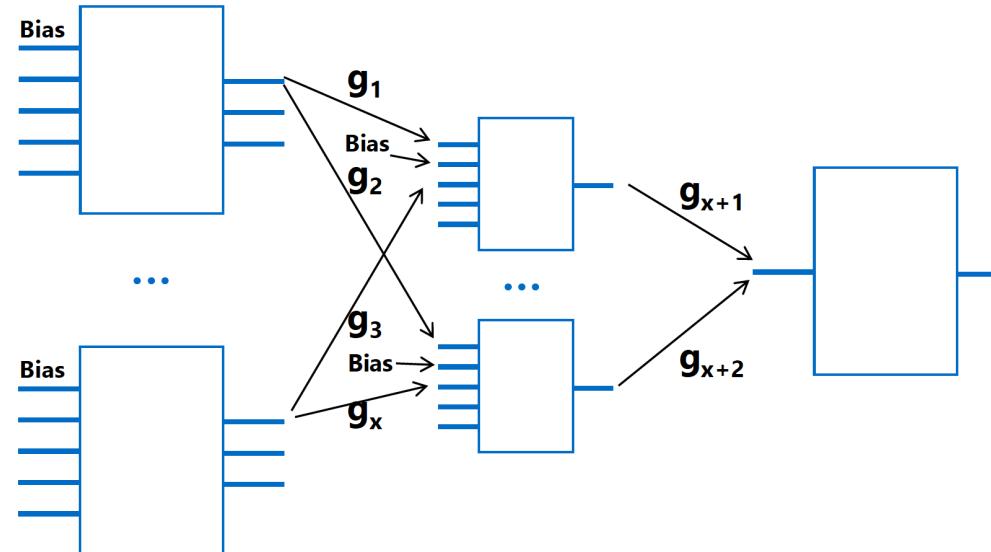
Cognitive Services



Cognitive Services

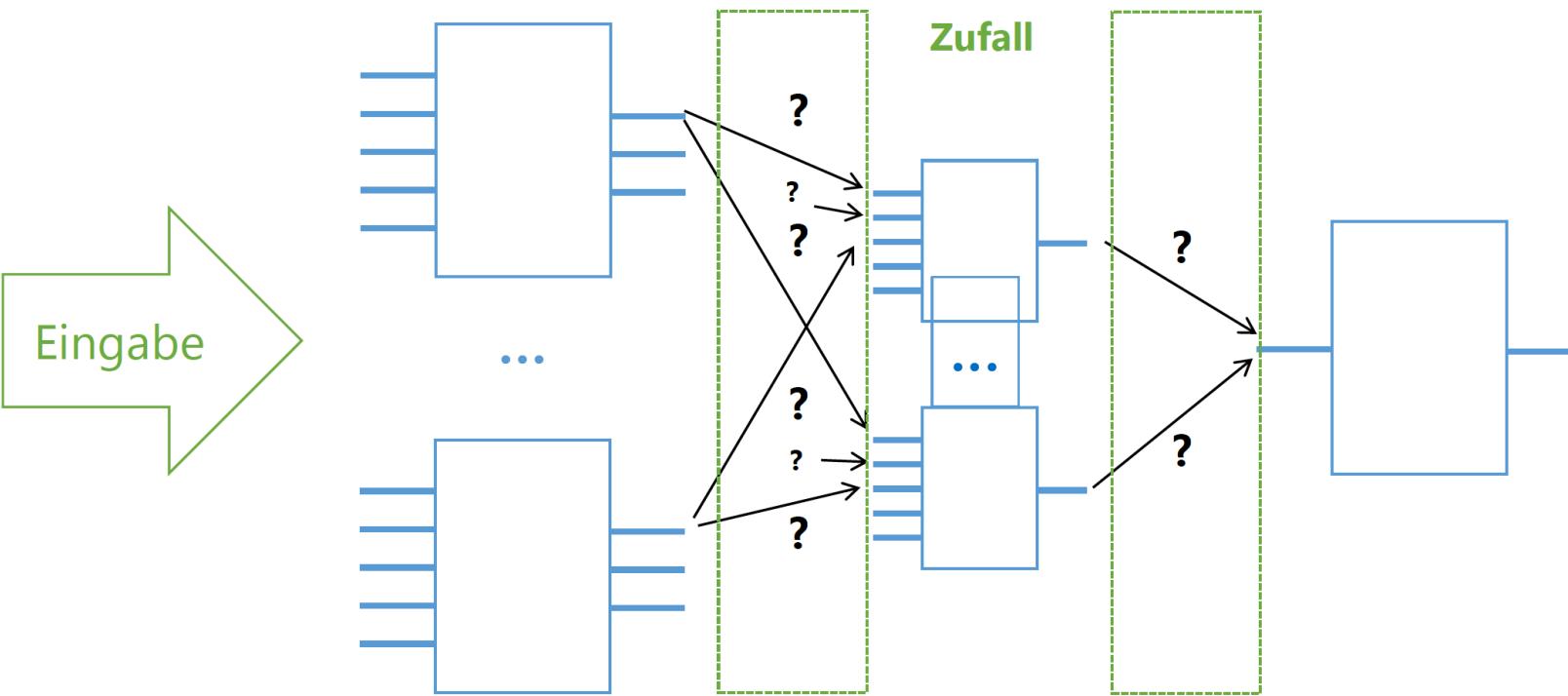


Single Layer



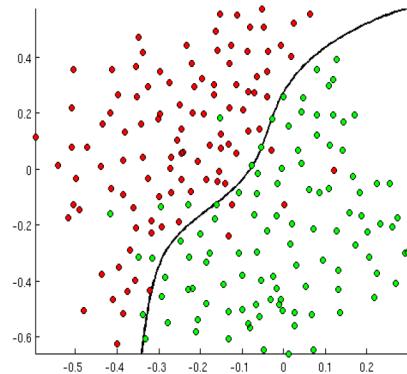
Multi-Layer

Cognitive Services

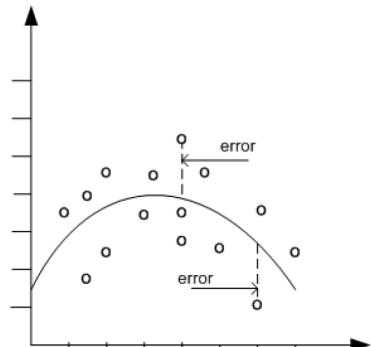


Cognitive Services

Klassifizierung



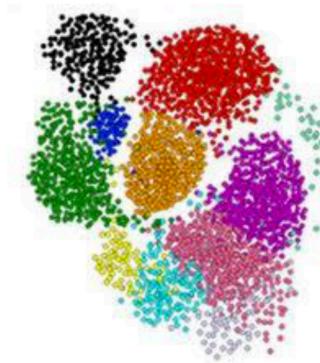
Regression



Assoziation
& Matchbox



Clustering



Cognitive Services

- **CPU**
- A central processing unit (CPU) is the electronic circuitry within a computer that carries out the instructions of a computer program by performing the basic arithmetic, logical, control and input/output (I/O) operations specified by the instructions.
- **GPU**
- A graphics processing unit (GPU), occasionally called visual processing unit (VPU), is a specialized electronic circuit designed to rapidly manipulate and alter memory to accelerate the creation of images in a frame buffer intended for output to a display device.
- **FPGA**
- A field-programmable gate array (FPGA) is an integrated circuit designed to be configured by a customer or a designer after manufacturing – hence "field-programmable".

Internal Microsoft Azure

Microsoft Translator V1.00.23400.11102

Data Source: War and Peace

Translate to: English

War and Peace

Author: Leo Tolstoy

Language: Russian

Pages: 1,440

Novel by Russian author Tolstoy, originally titled *Война и мир*, considered one of his finest literary achievements and a central work in world literature.

Azure Standard Server – 24-53N-1U

Type: 24-core 2.4 GHz Haswell

Model: 24-53245-NKD 1U

Peak Power/Unit: 300 Watts

Compute Capacity: 1.8 Tera-ops

Time Elapsed: 00:19.9

Pages Translated:  1440

Pages Per Second: 73

TRANSLATE

Azure FPGA Server – SV4-D5-1U

Type: 10 CPU cores + 4 FPGAs

Model: Stratix V D5-accelerator

Peak Power/Unit: 240 Watts

Compute Capacity: 7.9 Tera-ops

Time Elapsed: 00:02.6

Pages Translated:  1440

Pages Per Second: 558

TRANSLATE

Cognitive Services



Vision

Computer Vision | Emotion | Face | Video | Content

Moderator

Speech

Custom Recognition | Speaker Recognition | Speech

Language

Bing Spell Check | Language Understanding | Linguistic Analysis
| Text Analytics | Web Language Model | Translator Text and
Speech

Knowledge

Academic Knowledge | Entity Linking | QnA Maker Knowledge
Exploration | Recommendations

Search

Bing Auto Suggest | Bing Image Search | Bing News Search |
Bing Video Search | Bing Web Search

Story

Once Upon a time bla bla bla ...

Ale bla bla bla ...

Shoes bla bla bla ...

Sell via internet bla bla bla ...

Introduction to Recommendation

- Recommendation systems
 - Predict or rank the ratings a user may give to a lot of items
- Business goals of recommender systems
 - Increase product sales
 - Improve customer experience and satisfaction
- Technical goals of recommender systems
 - Relevance – recommend items that are relevant to a user
 - Novelty – recommend items that a user has not seen before
 - Serendipity – recommend items that are surprising to a user
 - Diversity – recommend dissimilar items to a user

Recommendations are everywhere

Screenshot of the Breuninger website showing a product page for a Parka GREENLAND jacket.

The page includes:

- Product image and variants.
- Product title: FJÄLLRÄVEN Parka GREENLAND.
- Color: NAVY.
- Price: € 349,99.
- Delivery information: Kostenfreie Lieferung per Click&Collect in Ihr Breuninger Haus in 1-2 Werktagen.
- Size options: M = 50, L = 52, XL = 54, XXL = 56.
- Product details tab.
- Material: 65% Polyester, 35% Baumwolle.
- Care instructions: Washable, Dry clean, Iron, Do not bleach.
- Share buttons: Teilen, Email, Facebook, Twitter, Print.
- Related products section: KUNDEN, DIE DIESEN ARTIKEL ANGEBEN HABEN, GEFÄLLT AUCH.
- Items shown: Marc O'Polo DENIM, BOSS Orange, SCOTCH & SODA, PAUL, FJÄLLRÄVEN, FJÄLLRÄVEN, Schmuddelweddha.

Screenshot of the Amazon.de website showing a product page for the LEGO Simpsons 71006 - Das Simpsons Haus.

The page includes:

- Product image: LEGO Simpsons 71006 - Das Simpsons Haus.
- Rating: ★★★★☆ - 19 Kundenrezensionen.
- Price: EUR 239,99.
- Description: Ein Lego-Aktionsgutschein zu gewinnen.
- Stock status: Nur noch 1 auf Lager.
- Delivery information: Lieferung 16.-20. Dez., wenn Sie Standardversand an der Kasse wählen. Siehe Details.
- Additional products section: Wird oft zusammen gekauft.
- Items shown: LEGO Simpsons 71016 - Das Simpsons Haus, LEGO Minifiguren The Simpsons Serie 2, LEGO 71005 Minifiguren, Unitedlabels - 0805380 - 71000 Minifiguren, Lego Simpsons Serie 2, Lego 40139 - Minifiguren, LEGO Ideas 21302 - The Big Bang Theory, LEGO 71009 Minifiguren, LEGO Technic 42006 - Drehleiter.

Recommendation Algorithms Offered by API

- Matrix Factorization-based Recommendation
 - Item Recommendations – infer latent feature vectors based on transactions and use the latent vectors to find similar items for an item
 - User Recommendations – provide personalized recommendations for user based on their transaction histories
 - Cold item placement – support cold items by incorporation item features

Recommendation Algorithms Offered by API

- Frequently Bought Together (FBT)
 - Recommend items that are likely to be purchased together in the same transaction with a particular item
- Item-item similarity is first computed based on training data
 - Transaction basket is considered
 - Items in the usage file with the same userId and timestamp are considered to be in the same transaction
 - Different similarity functions are provided
 - Co-occurrence
 - Lift
 - Jaccard similarity
- In scoring, a seed item is provided, and the most similar k items are returned

Basic Concepts in Recommendations API

- Project
 - A project is a container for your catalog data, usage data and your recommendations builds
- Build
 - You can create different builds under a project using different recommendation algorithms and settings with the same catalog and usage data
 - Three different kinds of builds in Recommendations API
 - Rank build
 - Recommendation build
 - Frequently Bought Together (FBT) build

Model and Build in Recommendations API

The screenshot shows the Microsoft Cognitive Services Recommendations UI-(Beta) interface. The title bar indicates the project is "Demo01". The main navigation bar includes links for "Projects", "Sample Code", "Documentation", and "API Reference". A "Logged in" user icon is also present.

The main content area is titled "Demo01" and displays a Model ID: 7360cda1-4607-425f-aa77-6605173bae6d. It is divided into three sections:

- Step 1: Add a catalog file (schema)**: Contains a "BROWSE FOR FILE..." button. Below it, a message says "Catalog uploaded".
- Step 2: Add usage files (schema)**: Contains a "BROWSE FOR FILE..." button. Below it, a message says "Uploaded files".
- Step 3: Create a build**: Contains a "NEW BUILD" button. Below it, a table lists two builds:

ACTIVE BUILD	BUILD ID	TYPE	STATUS	DATE CREATED	SCORE
●	1591030	Recommendation	Success	2.12.2016, 11:03:04	SCORE
	1591028	Rank	Success	2.12.2016, 10:56:13	FEATURES

At the bottom left, there is a link to "TCLCartRemovePGProd.csv" and a trash bin icon. The bottom right corner features the SQL Saturday logo.

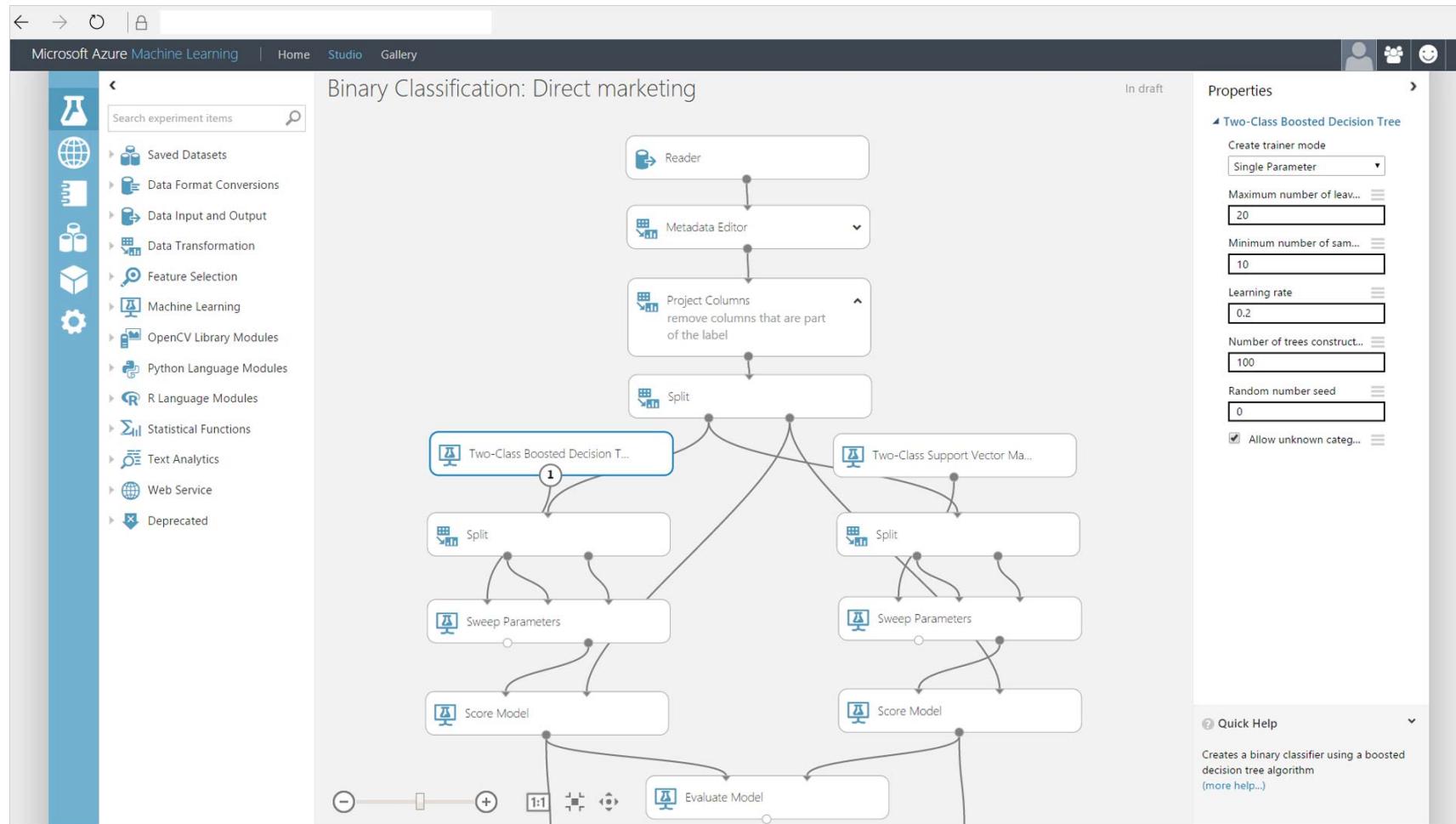
How to Use Recommendations API

- Http calls
 - All requests must be made in http calls
- Simple GUI tools are provided
 - Supports creating model, uploading data, training builds and test recommendation
- C# sample code

Use CS Recommendations API

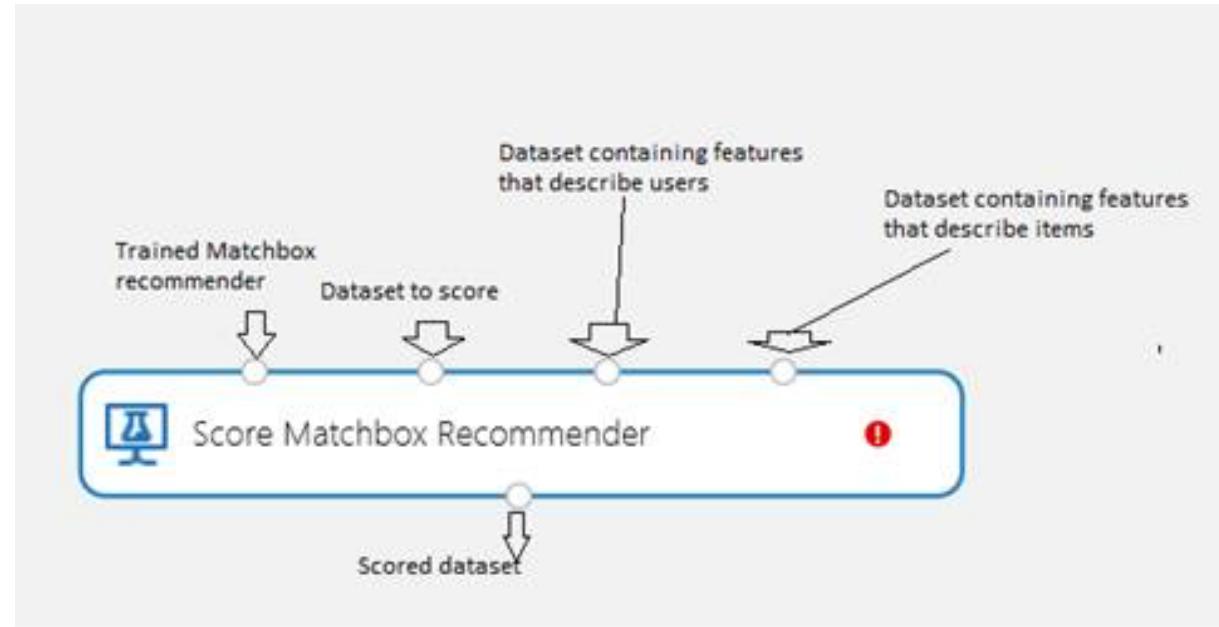
Demo ...

Azure ML



Matchbox Recommendation Algorithms Offered

- Rating Prediction
- Item Recommendation
- Related Users
- Related Items



How to Use Recommendations API

- Http calls
 - All requests must be made in http calls

Use Azure ML Recommendations

Demo ...

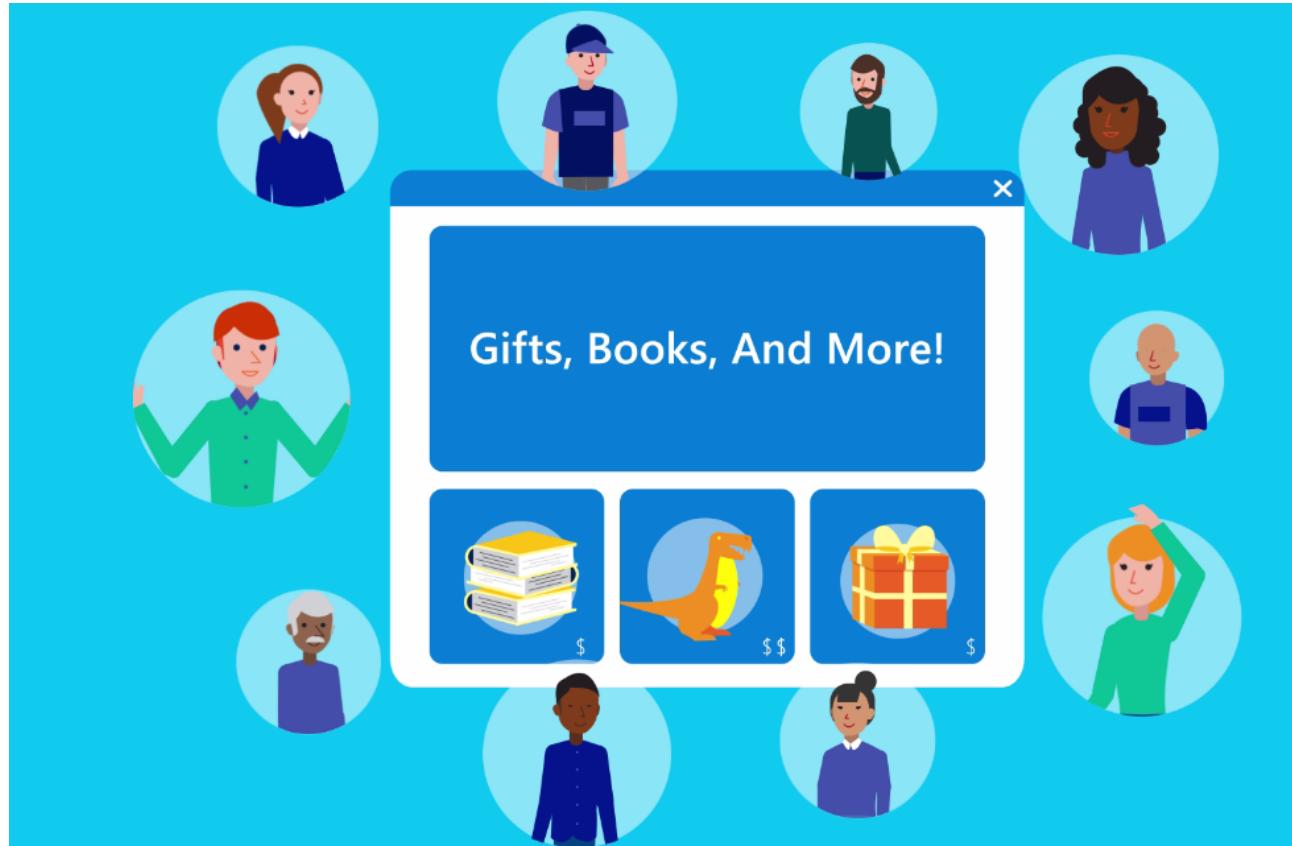
Review

	Azure ML					CS Recommendation API				
Knowledge										
Implementation time										
Customizing										
Scaling										
Pricing										

Next steps

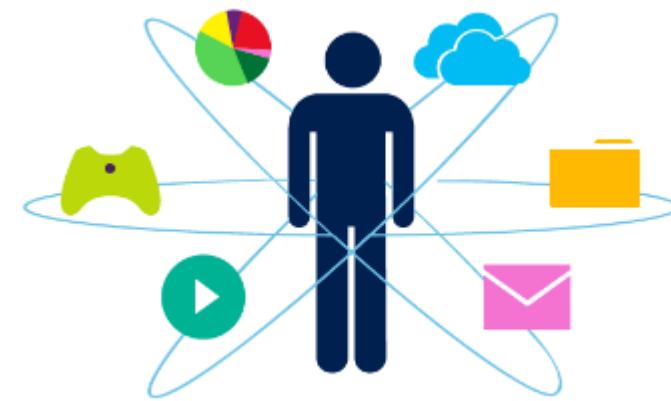
- A/B test
- personalized mailings
- personalized landing pages
- ...

Questions? Questions!



Thank you

- @SQL_Alex
- <https://www.consulting-bi.de>
- a.klein@consulting-bi.de



Cognitiv Services Pricing

Recommendations API	Free	10K calls per month	Free
	S1 Standard	100K calls per month	\$75 per month, overage at \$0.75 per 1K calls
	S2 Standard	1M calls per month	\$500 per month, overage at \$0.75 per 1K calls
	S3 Standard	10M calls per month, overage at \$0.75 per 1K calls	\$2,500 per month, overage at \$0.75 per 1K calls
	S4 Standard	50M calls per month	\$5,000 per month, overage at \$0.75 per 1K calls

Azure ML Pricing

REGION: West Europe FEATURE: API Usage - New Web Service

TIER: Standard S1

Includes 100,000 API transactions and 25 API compute hours.

Standard S1 base price = €84.33/MO

1000000 Transaction overages × €0.42 Per 1,000 = €421.65/MO

0 Compute hour overages × €1.69 Per hour = €0.00/MO

Sub-total €505.98/MO