



# How-to: Bilderkennung mit SAP-HANA, TensorFlow und Jupyter

Luzern, 05.05.2020  
Alexander, Ramm

**pure**consulting.

# Wer bin ich?



**Alexander Ramm**  
**Senior Consultant**

**Twitter:** [@alexander\\_ramm](https://twitter.com/alexander_ramm)

**SAP:** [people.sap.com/pureconsulting\\_alexander.ramm](https://people.sap.com/pureconsulting_alexander.ramm)

# Warum diese Präsentation

- Wie aufwendig ist es eine Eigene Machine Learning Plattform aufzusetzen
- Wie kann ich SAP HANA mit externen Tools verbinden und die Plattform so umfassender einfacher verwenden
- Kann ich die ML-Fähigkeiten von SAP HANA selber erweitern und wie aufwendig ist dies

# Was wird heute gezeigt?

## Integration von SAP Hana mit externen Tools im Case Maschinelles Lernen

- **101 Machine Learning**
- **Erweiterungsmöglichkeiten der SAP HANA**
- **Jupyter Notebook**
- **TensorFlow**
- **Beispiel**

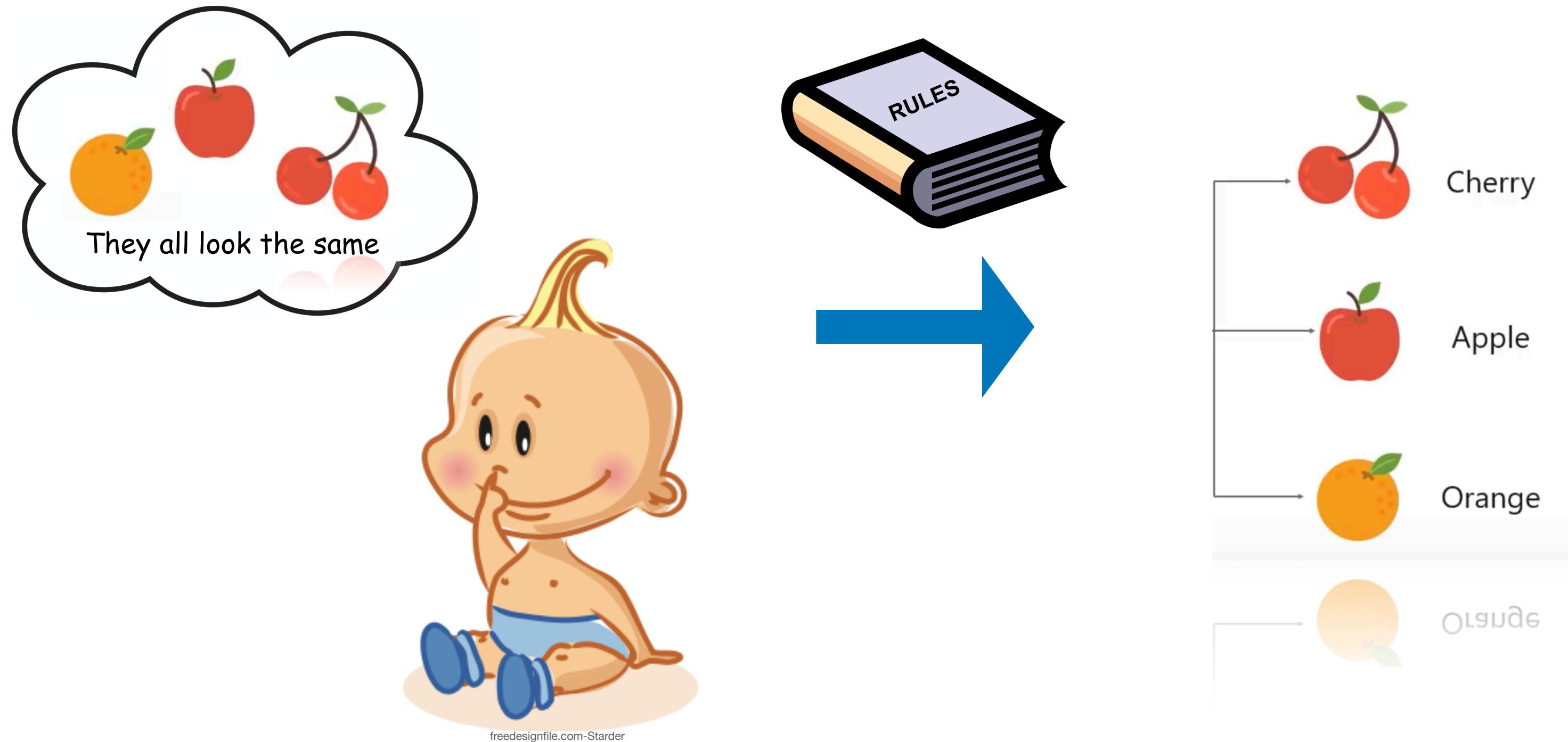
# 101 - Machine Learning

# Was ist Machine Learning

**Maschinelles Lernen** ist ein Oberbegriff für die „künstliche“ Generierung von Wissen aus Erfahrung: Ein künstliches System lernt aus Beispielen und kann diese nach Beendigung der Lernphase verallgemeinern.

— de.wikipedia.org

# Was ist Machine Learning



# Was ist Machine Learning

Beim Machine Learning bezeichnet man unsere Regeln als Model und diese werden erlernt und nicht gecodeed.

Klassisches Programm:



```
if (siehtAusWieApfel)  
    isLabel: "Apfel"
```



# Was ist Machine Learning

Model und Inputdaten treffen eine Vorhersage.



# Was ist Machine Learning

Das Model wird mit speziellen Algorithmen erzeugt welches Eingabedaten verallgemeinert.



# Was ist Machine Learning - Hauptarten

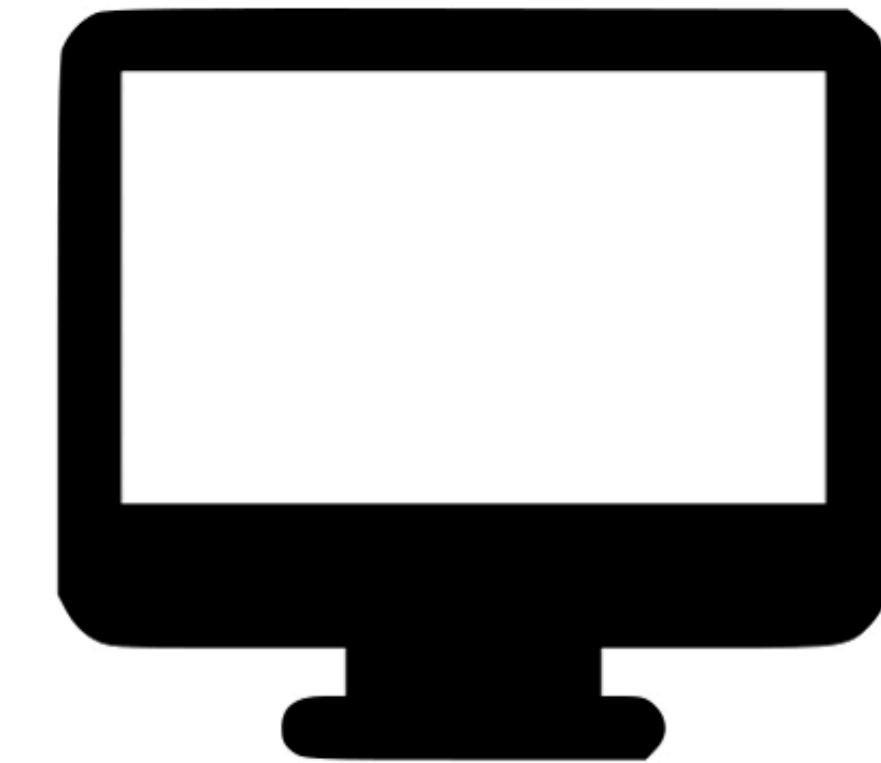
- **Supervised Learning (Überwachtes Lernen)**
  - **Regressionsproblem**
  - **Klassifikationsproblem**
- **Unsupervised Learning (Unüberwachtes Lernen)**
  - **Segmentierung**
  - **Reinforcement Learning (Bestärkendes Lernen)**
  - **Künstliche Neuronale Netze**
  - **Deep Learning**

# Supervised Learning

Wir fungieren als “Lehrer”.



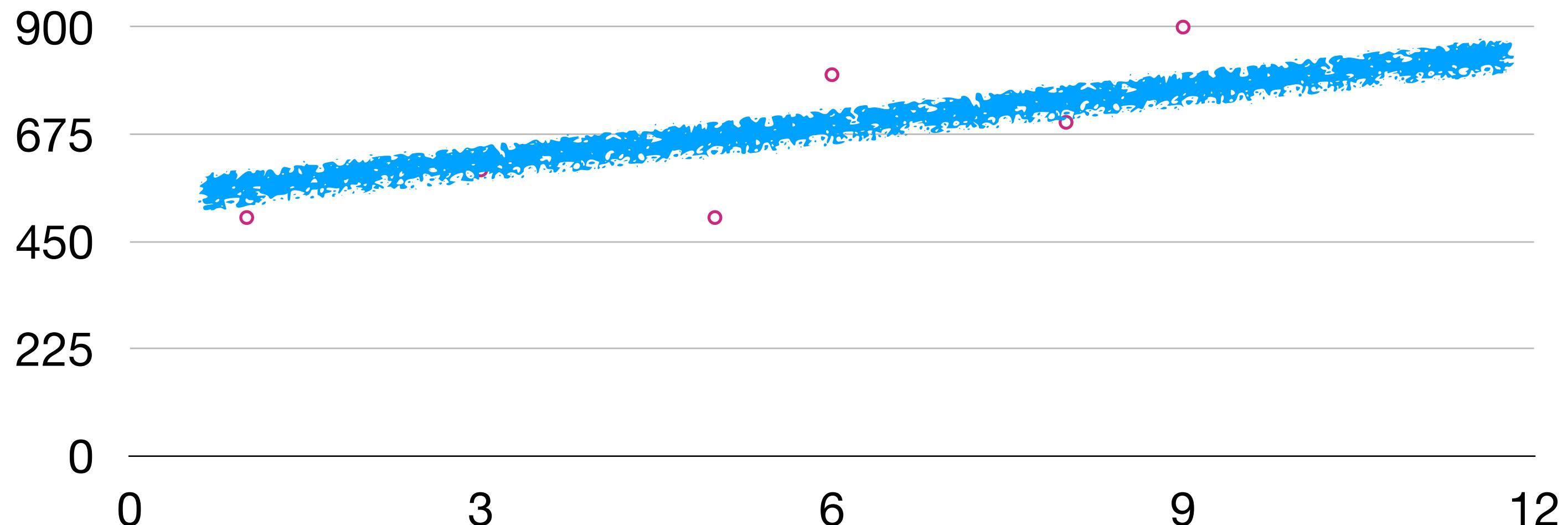
Wir kennen die Frage und Antwort (Labeled data).



# Supervised Learning

- **Regression**

Verwendung der Regressionsanalyse (z.B. Median). Output hat viele Werte. Wird zum Beispiel für Vorhersagen verwendet.



## Beispiele:

Vorhersage von Preisen, Unfällen

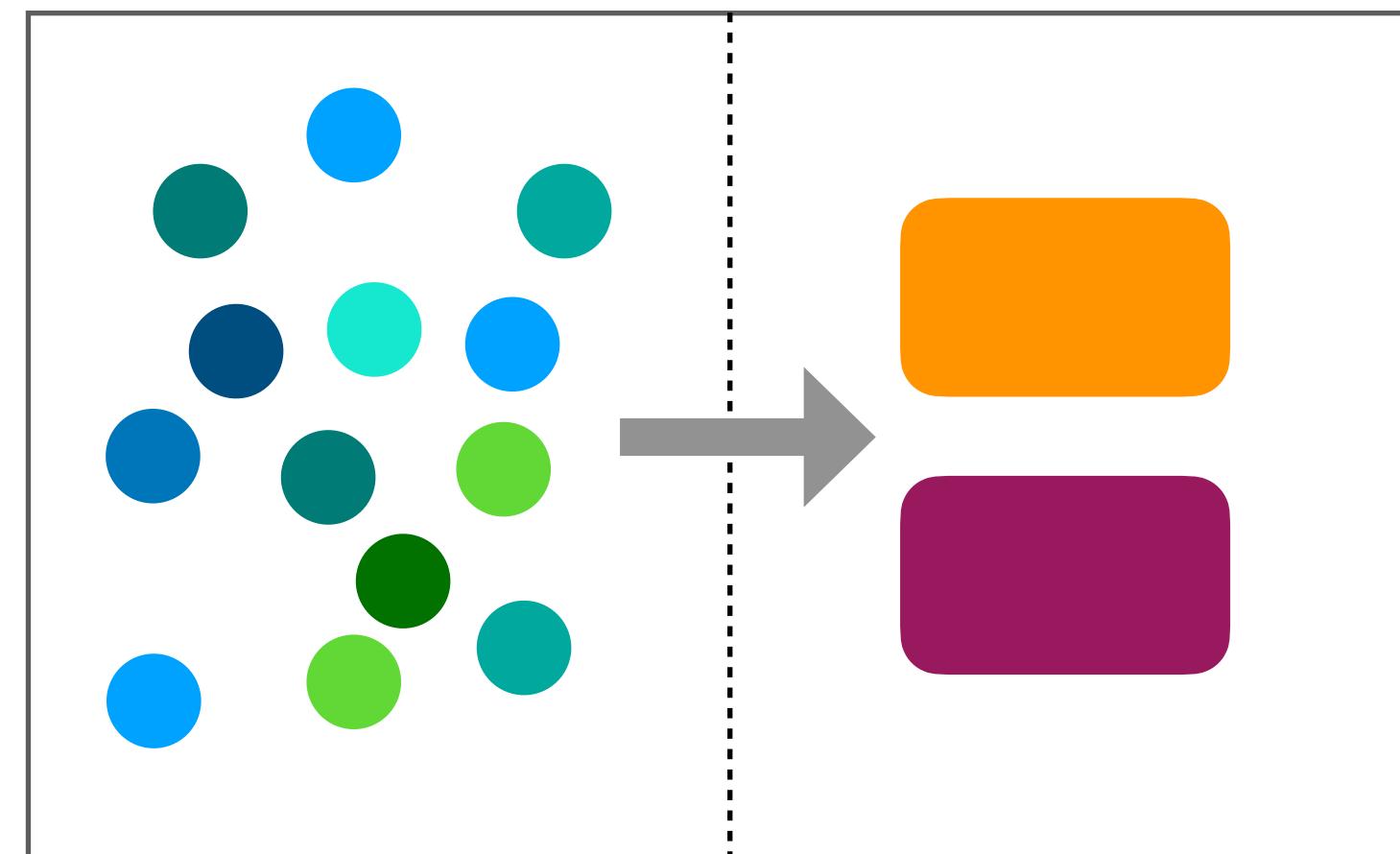
Erstellung von Prognosen:

- COVID-19 Infektionsmodelle
- Einkaufs-/Absatzprognosen

# Supervised Learning

- **Klassifikation**

Verhältnismässig hat der Output wenig Werte (Ja / Nein)



**Beispiele:**

Erkennung von Facebook-Posts/Twitter:  
Positive/Negative Einstellung zum Produkt

Erkennung von SPAM-Emails

# Unsupervised Learning

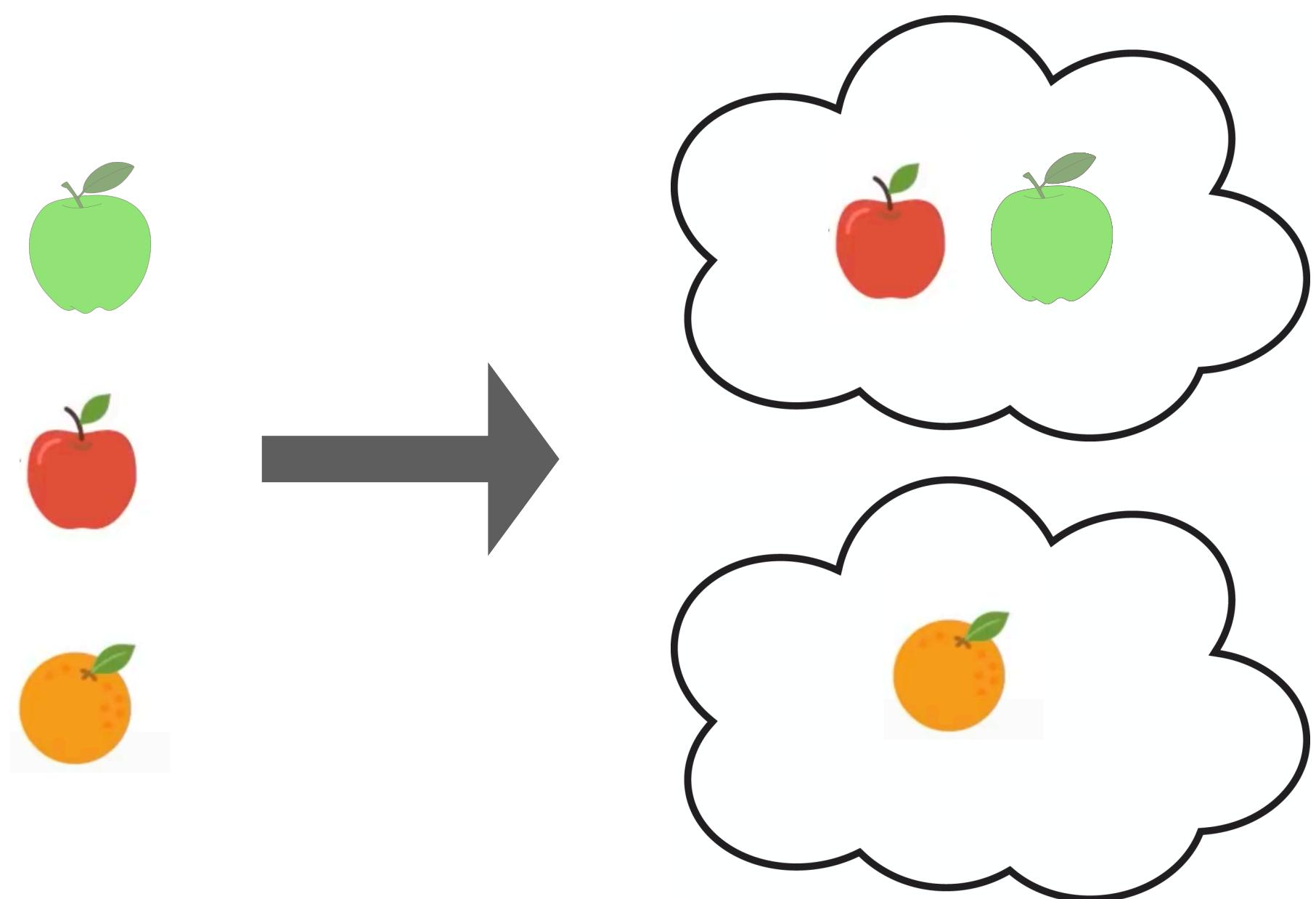
Wir können die Daten nicht beschreiben oder kennen sie nicht genau



# Unsupervised Learning

- **Segmentierung**

Es wird versucht Muster (Cluster) in den Daten zu erkennen.



**Beispiele:**

Spotify -> Musikempfehlung

Amazon -> Produktempfehlungen

# Reinforcement Learning

Es wird selbstständig eine Strategie entwickelt.

Bewertung der Ergebnisse und Optimierung des Models durch Belohnung.



# Künstliche Neuronale Netzwerke (KNN)

“Nachbildung” von natürlichen neuronalen Netzstrukturen

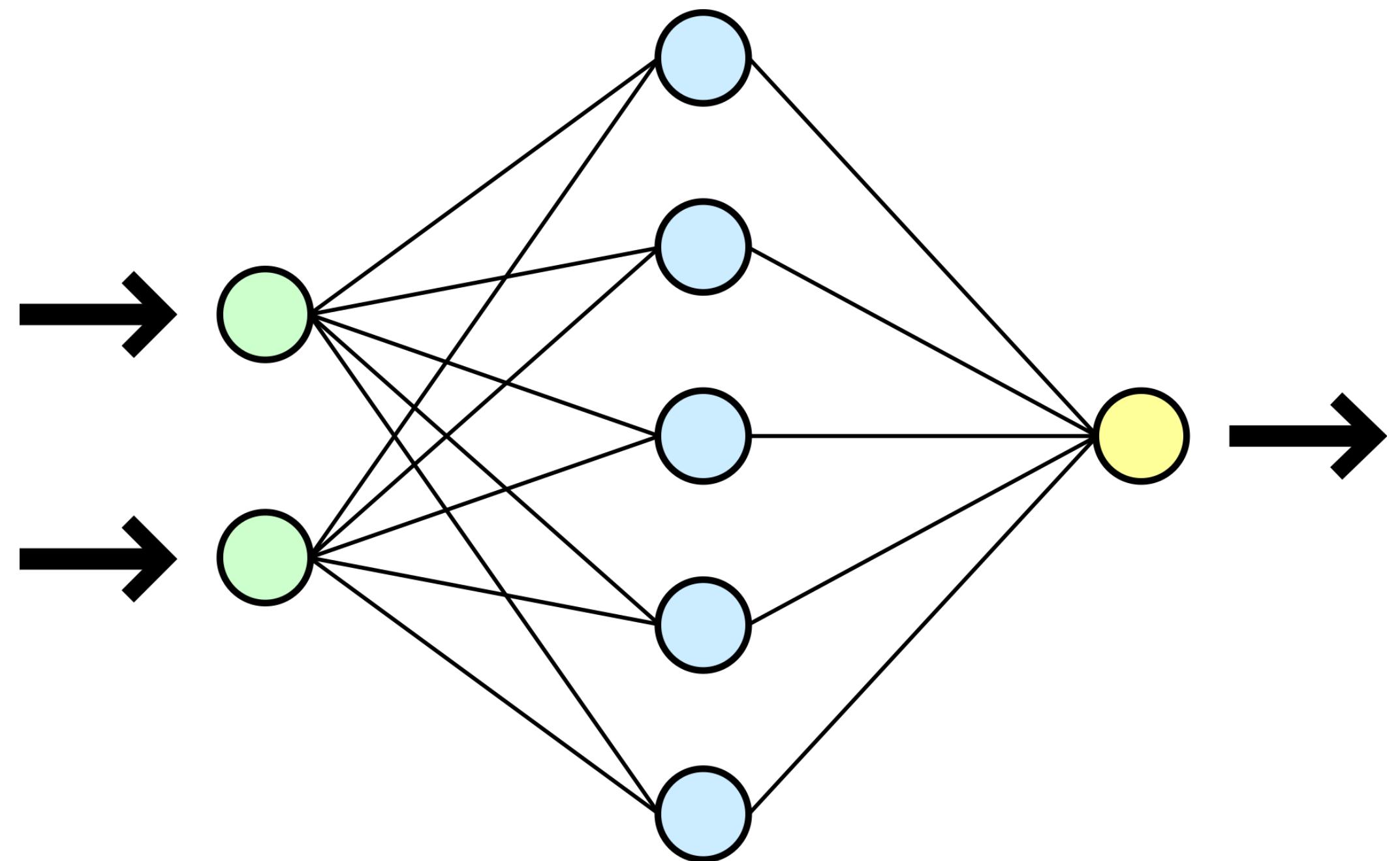


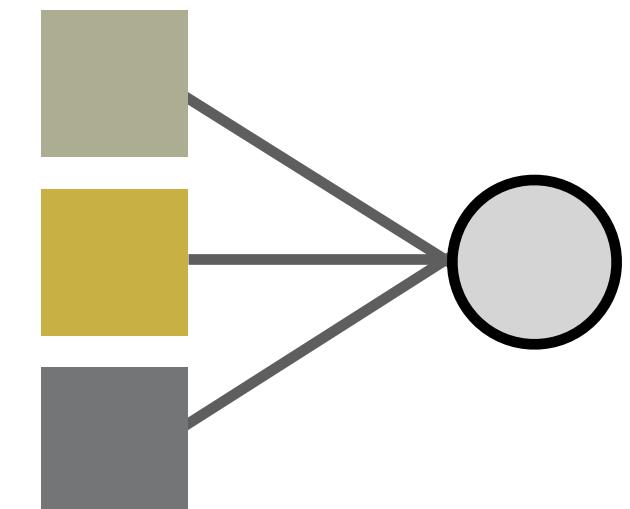
Bild Wikipedia

# Künstliche Neuronale Netzwerke (KNN)

Neuronen haben eine Input.



Photo by Max Baskakov on Unsplash



# Künstliche Neuronale Netzwerke (KNN)

Neuronen haben Parameter. Schwellenwert und Gewichtung.

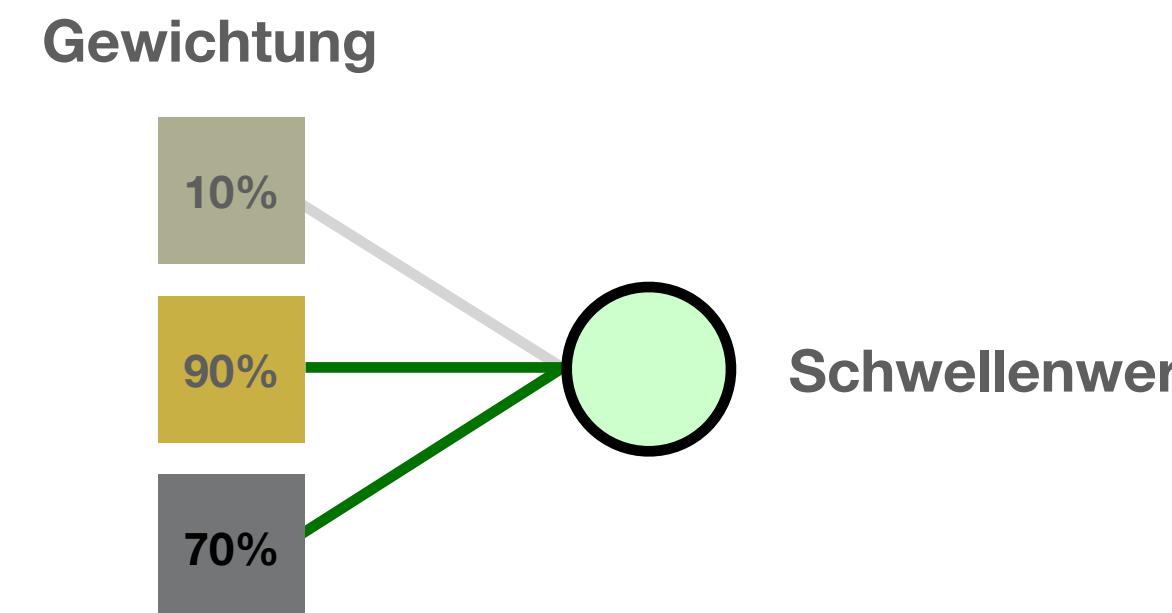
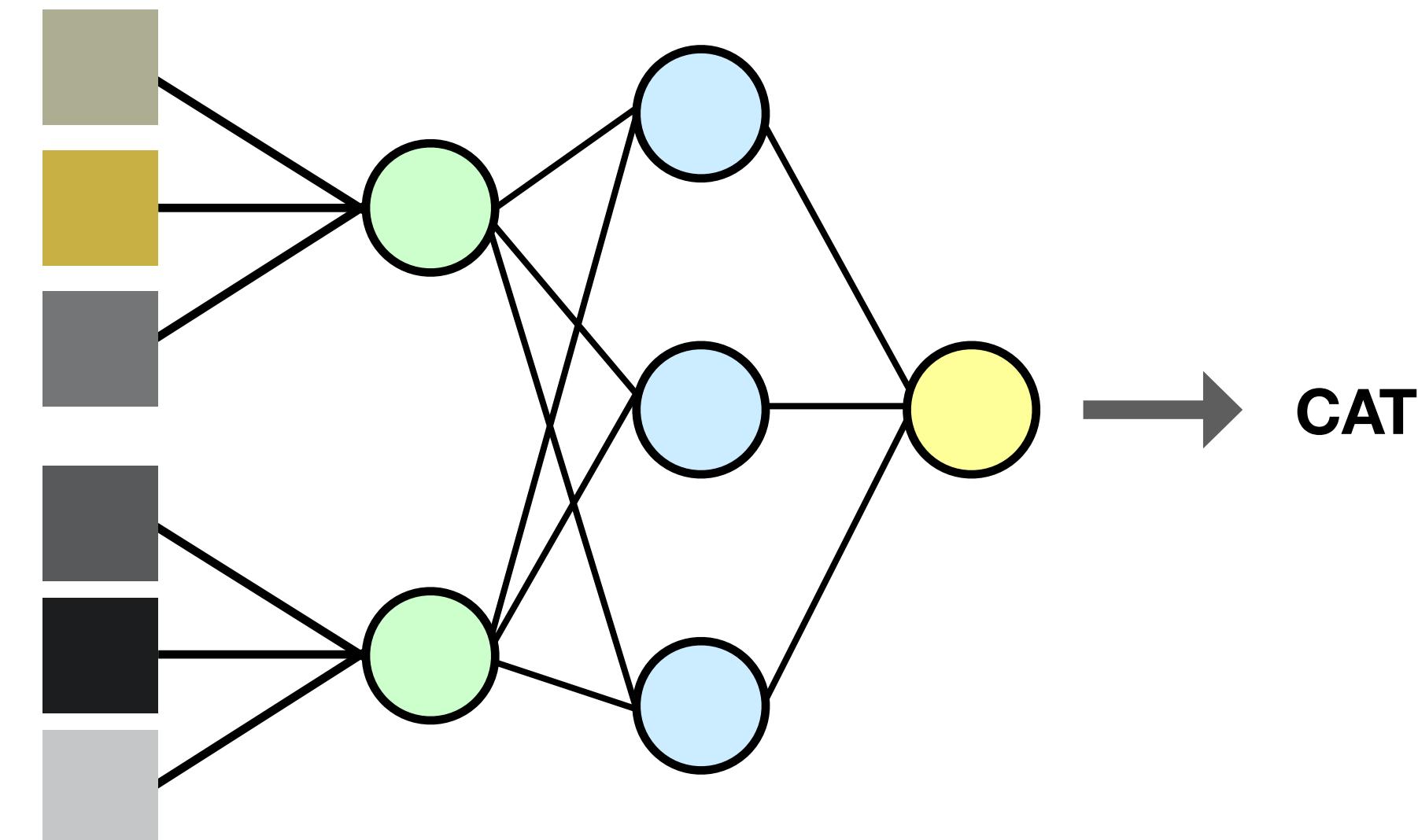


Photo by Max Baskakov on Unsplash

# Künstliche Neuronale Netzwerke (KNN)

Diese Parameter werden durch das “Lernen” parametrisiert und als “Model” gespeichert.



## Lernphase:

- Viele Input-Daten
- Parametrierung der Gewichtung und Schwellenwerte
- Löschen existierender Verbindungen
- Entwicklung neuer Verbindungen

# Deep Learning

Kombination mehrerer Verfahren/Algorithmen. Wir kennen meist nur den 1. Layer (Input) und den letzten Layer (Output). Layer dazwischen werden als Hiddenlayer bezeichnet.

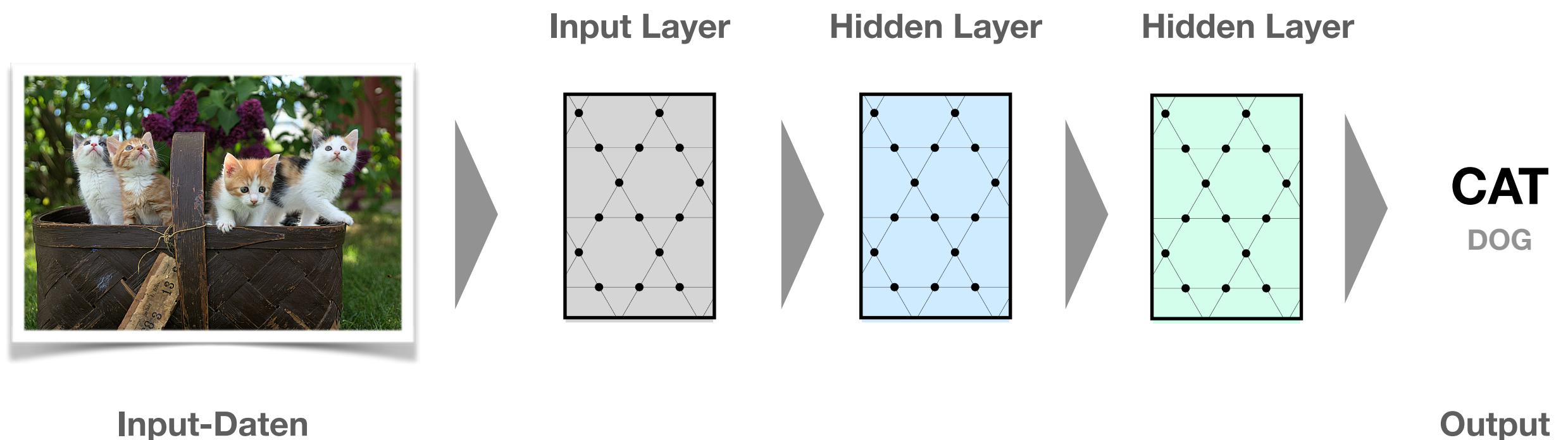


Photo by Jari Hytönen on Unsplash

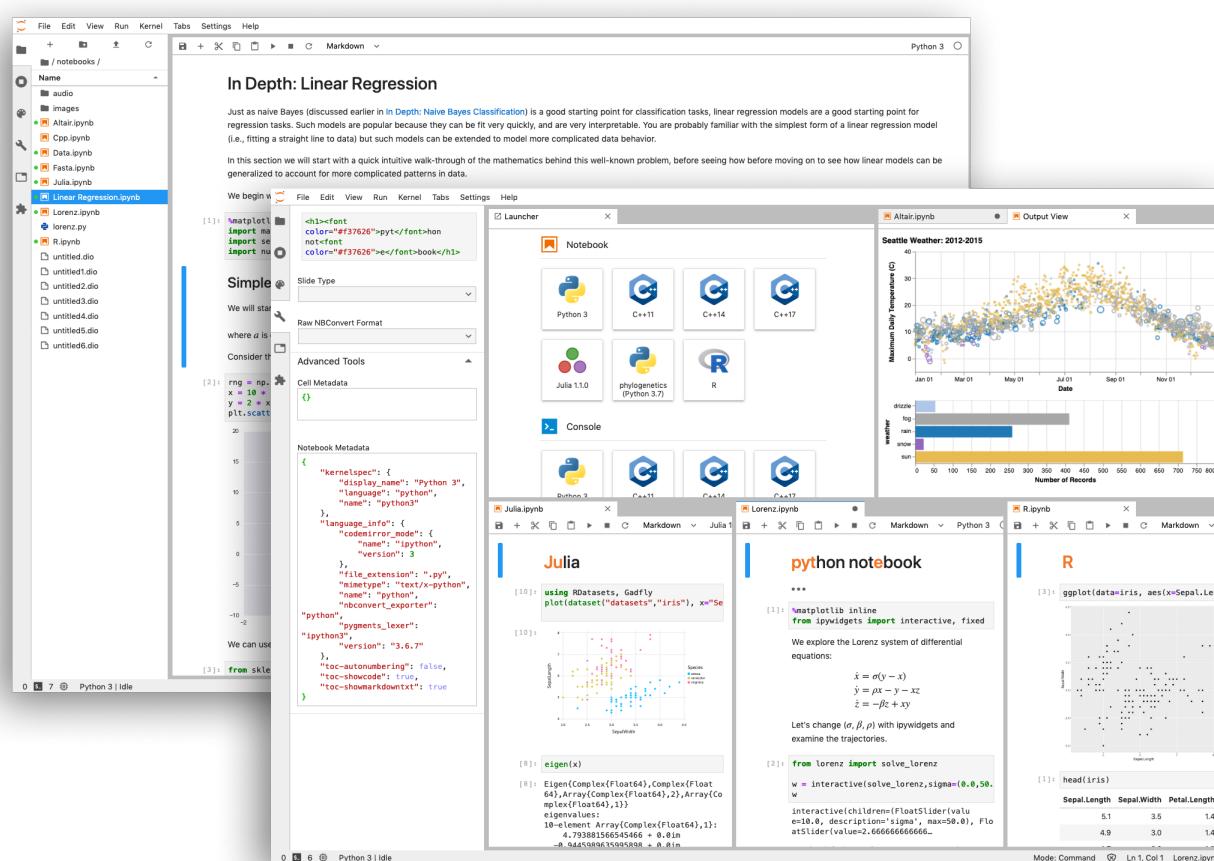
# Überblick

- **101 Machine Learning**
- Erweiterungsmöglichkeiten der SAP HANA
- Jupyter Notebook
- TensorFlow
- Beispiel

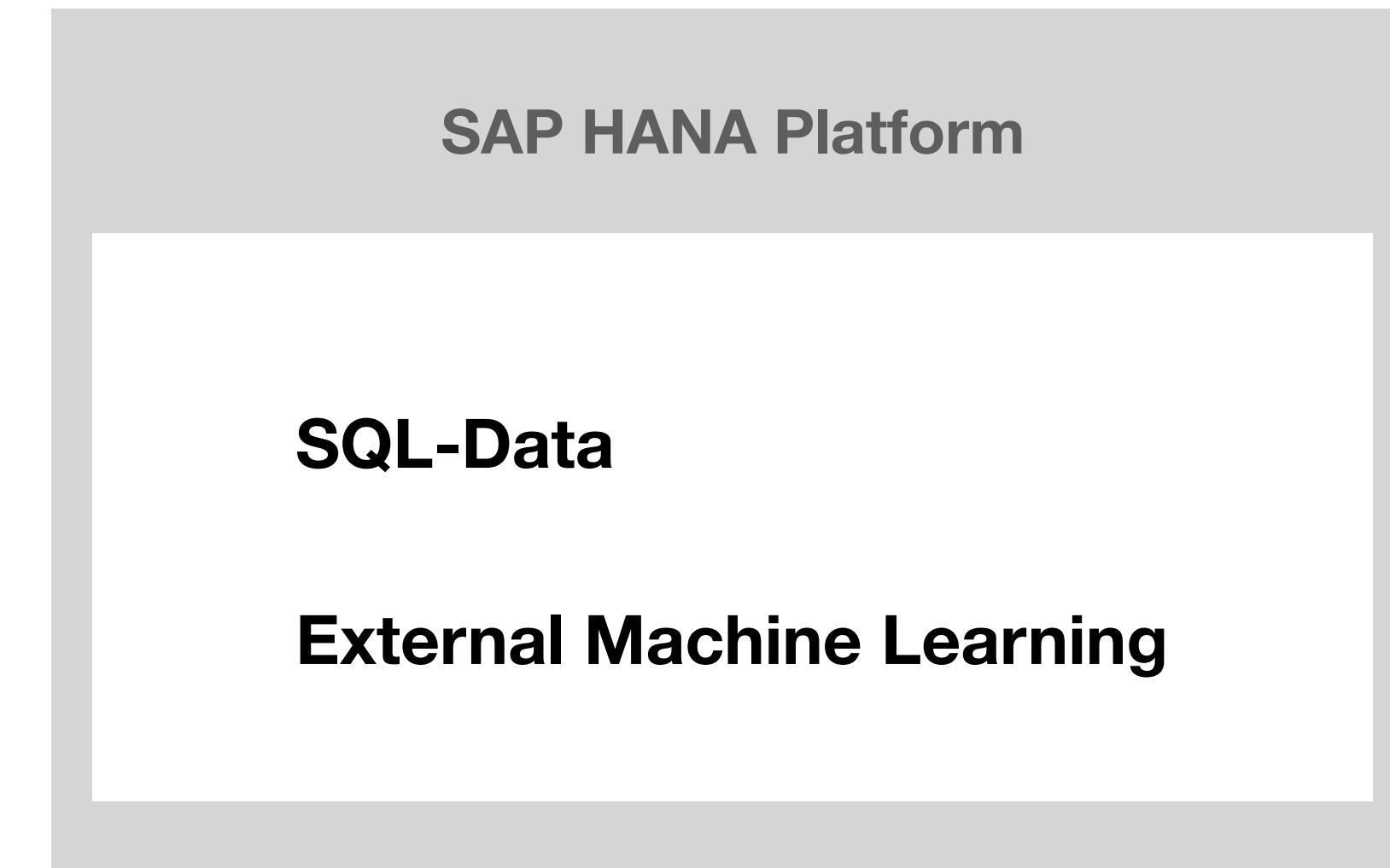
# SAP Hana erweitern

# Erweiterungsmöglichkeiten der SAP HANA

SAP HANA bietet unterschiedliche Möglichkeiten um externe Tools zu Integrieren



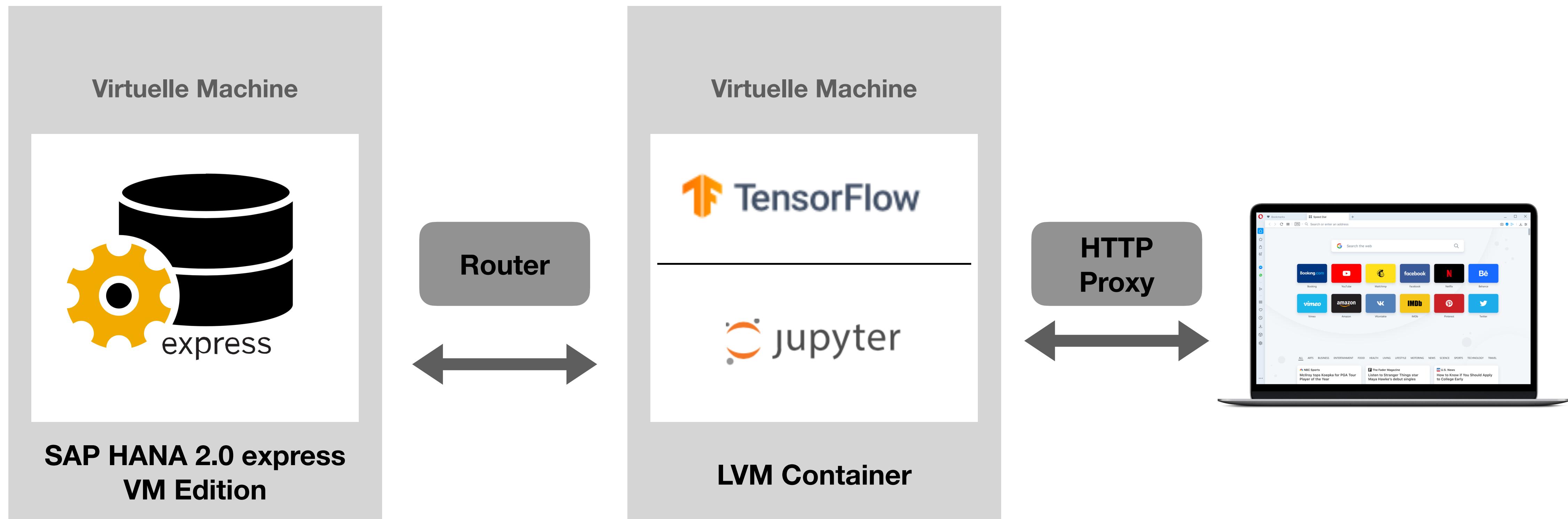
ODBC  
Python



EML call TensorFlow

# Demo/Schulungsinstanz

Getrennte Virtuelle Machine je Komponente

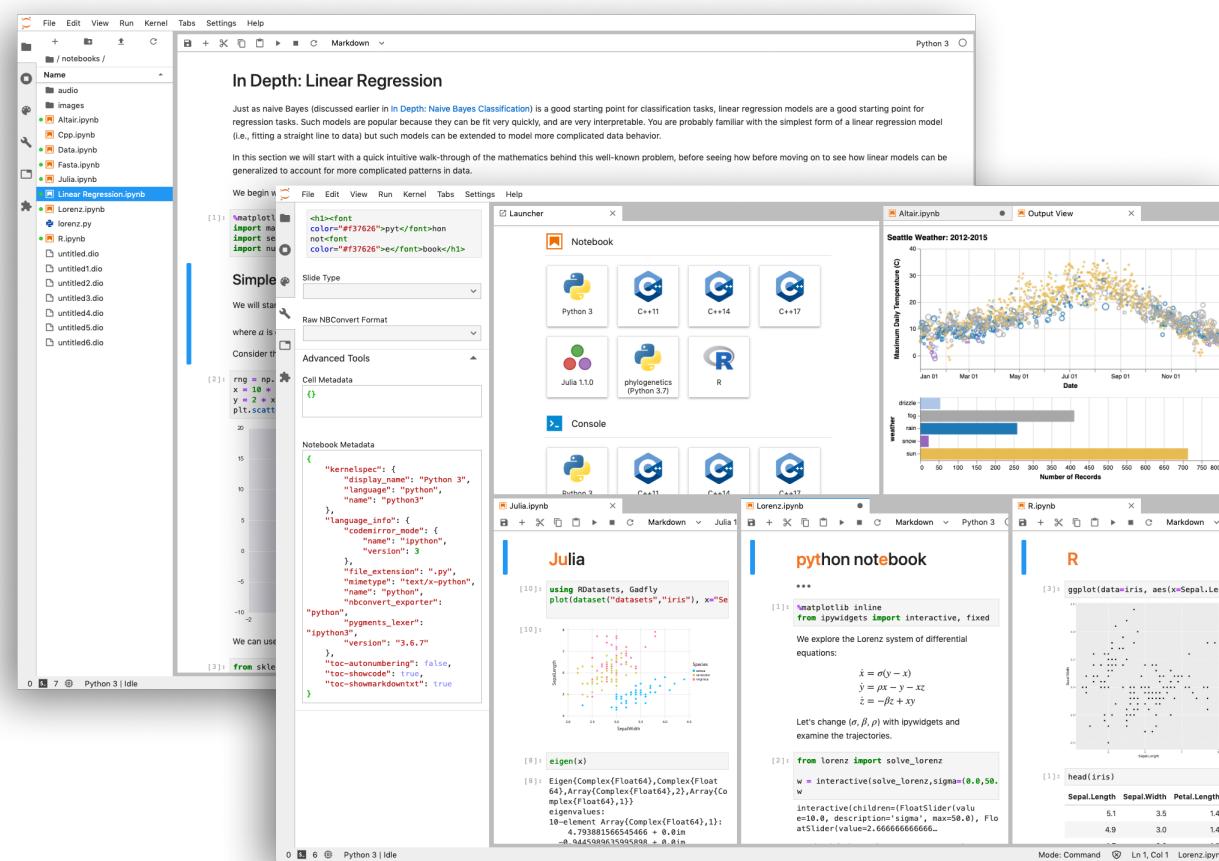


# Überblick

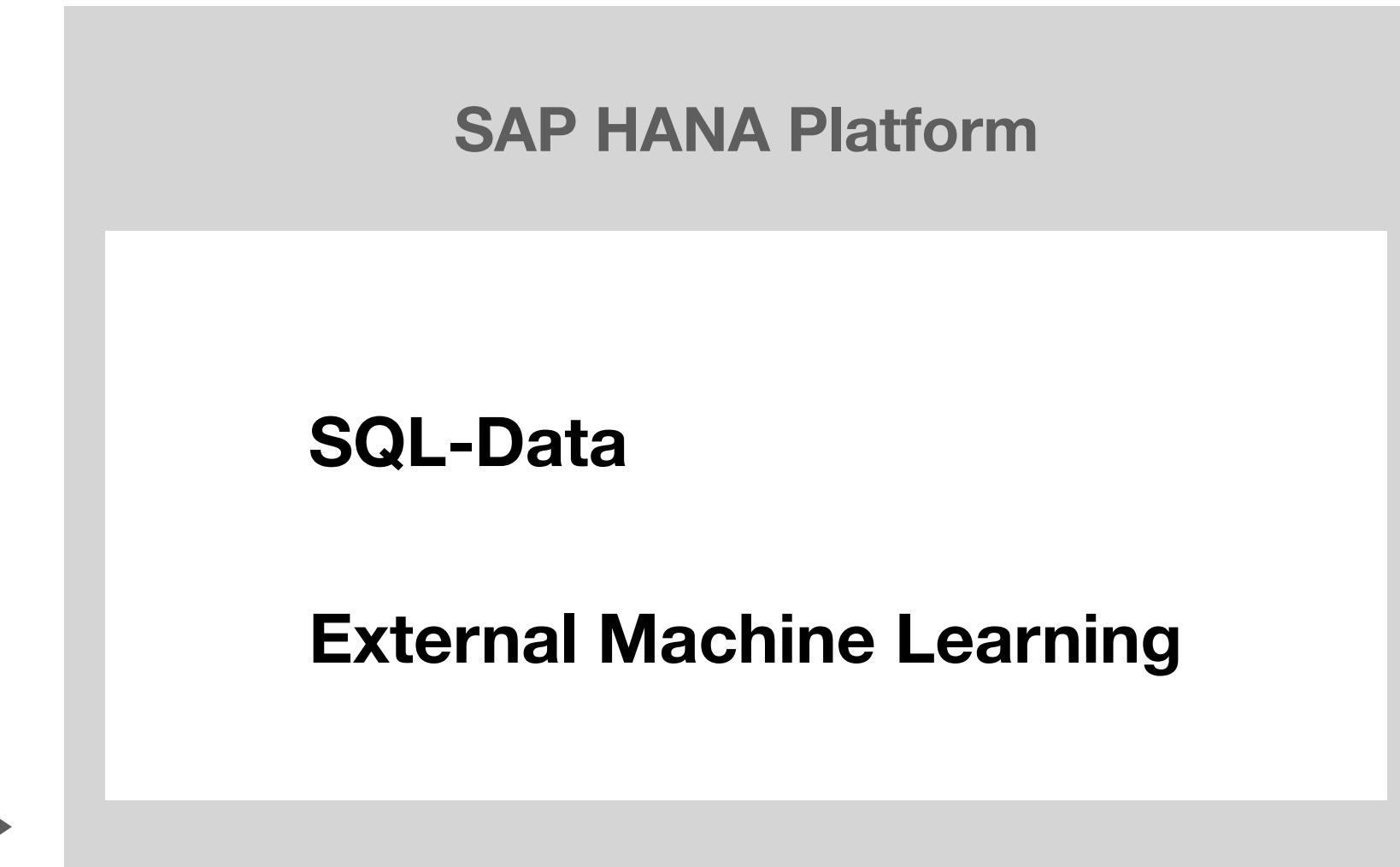
- **101 Machine Learning**
- **Erweiterungsmöglichkeiten der SAP HANA**
- Jupyter Notebook
- TensorFlow
- Beispiel

# Erweiterungsmöglichkeiten der SAP HANA

SAP HANA bietet unterschiedliche Möglichkeiten um externe Tools zu Integrieren



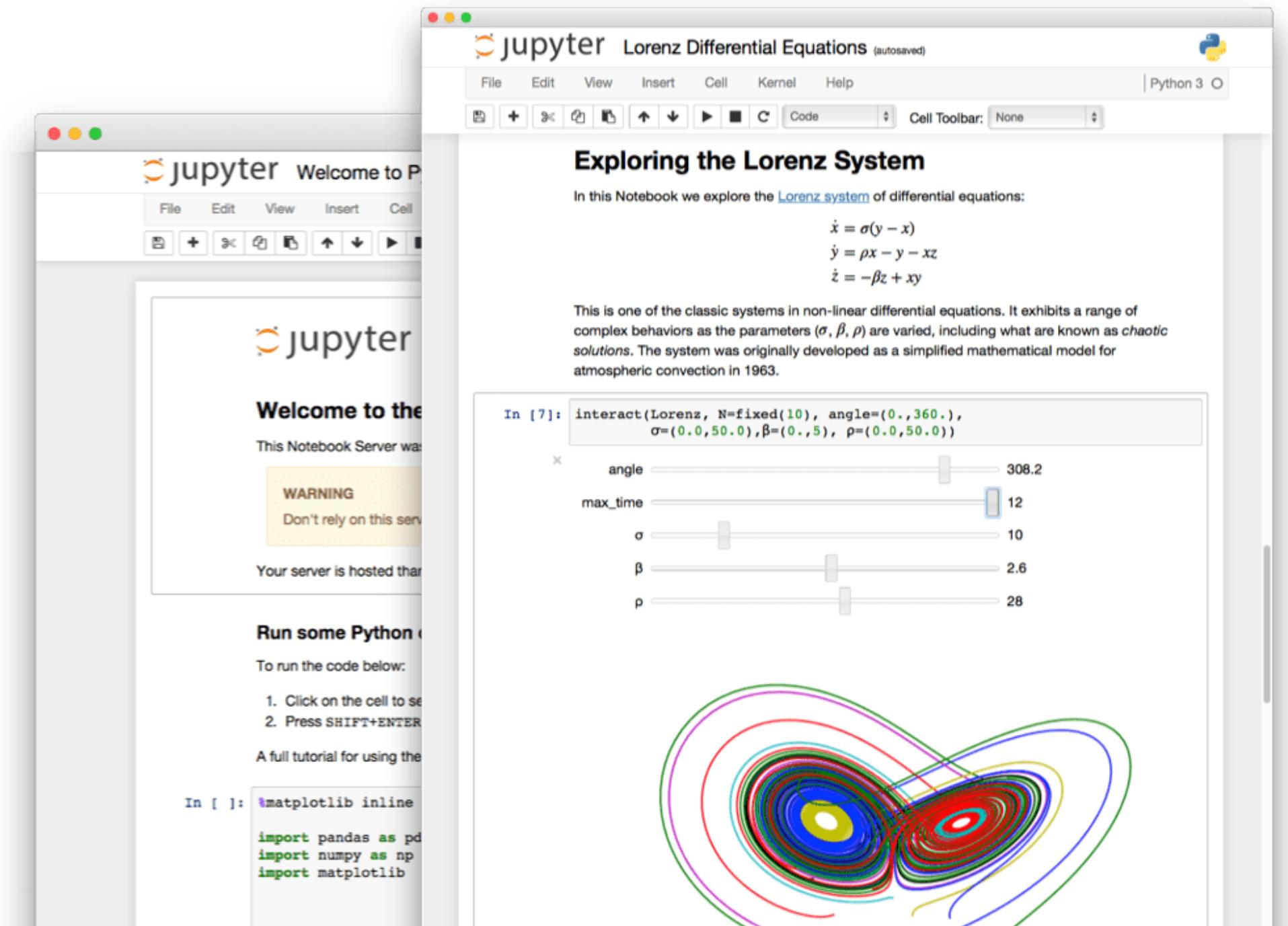
ODBC  
Python



EML call TensorFlow

# Jupyter Notebook

OpenSource WebApplikation um Dokumente zu erzeugen welche Livecode enthalten und an Datenquellen anschlossen sind.

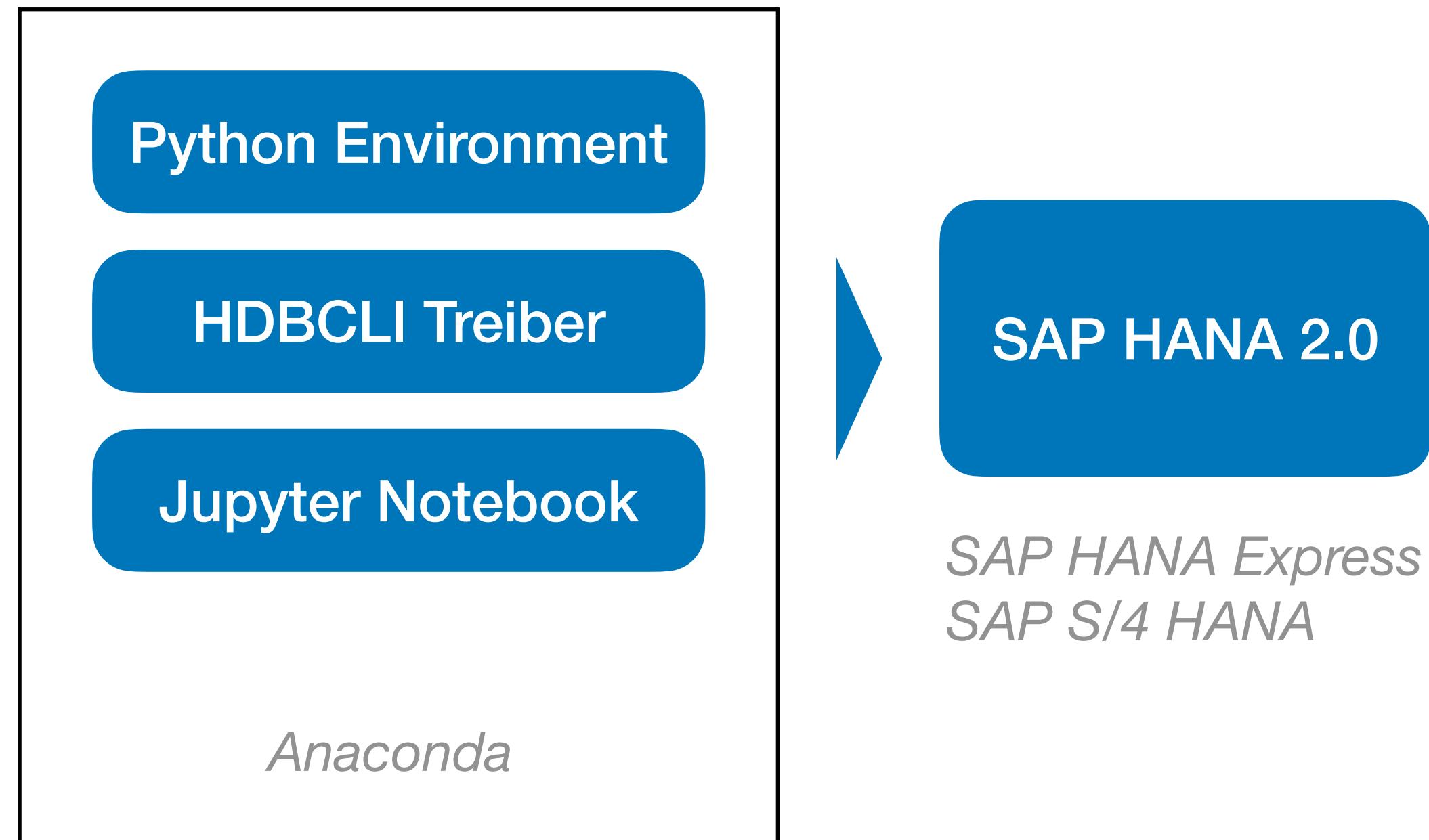


## Feature:

- Unterstützung vieler Sprachen (**Python, JS, R, ...**)
- Shareable (Notebooks können geteilt werden, [nbviewer.jupyter.org](http://nbviewer.jupyter.org))
- Visualisierung von Inhalten (Graphen, Diagramme, 3D)
- Integration (z.B. SAP HANA, Tensorflow, ...)
- HANA Integration -> Daten bleiben auf der DB

# Jupyter Notebook

## • Installation und Anbindung an SAP HANA



### Schritte:

- Python Environment und Jupyter aufsetzen. Am einfachsten mit Anaconda.
- HDBCLI (Python Treiber für HANA) installieren
- Mit HANA verbinden

# Jupyter Notebook

## • Installation und Anbindung an SAP HANA

```
!pip install hdbcli > /dev/null 2>&1
!if [ ! -f hana_ml-1.0.7.tar.gz ]; then wget
https://s3.amazonaws.com/sap_s3/
hana_ml-1.0.7.tar.gz > /dev/null 2>&1; fi
!pip install hana_ml-1.0.7.tar.gz > /dev/
null 2>&1
!pip show hana
```

### Probleme:

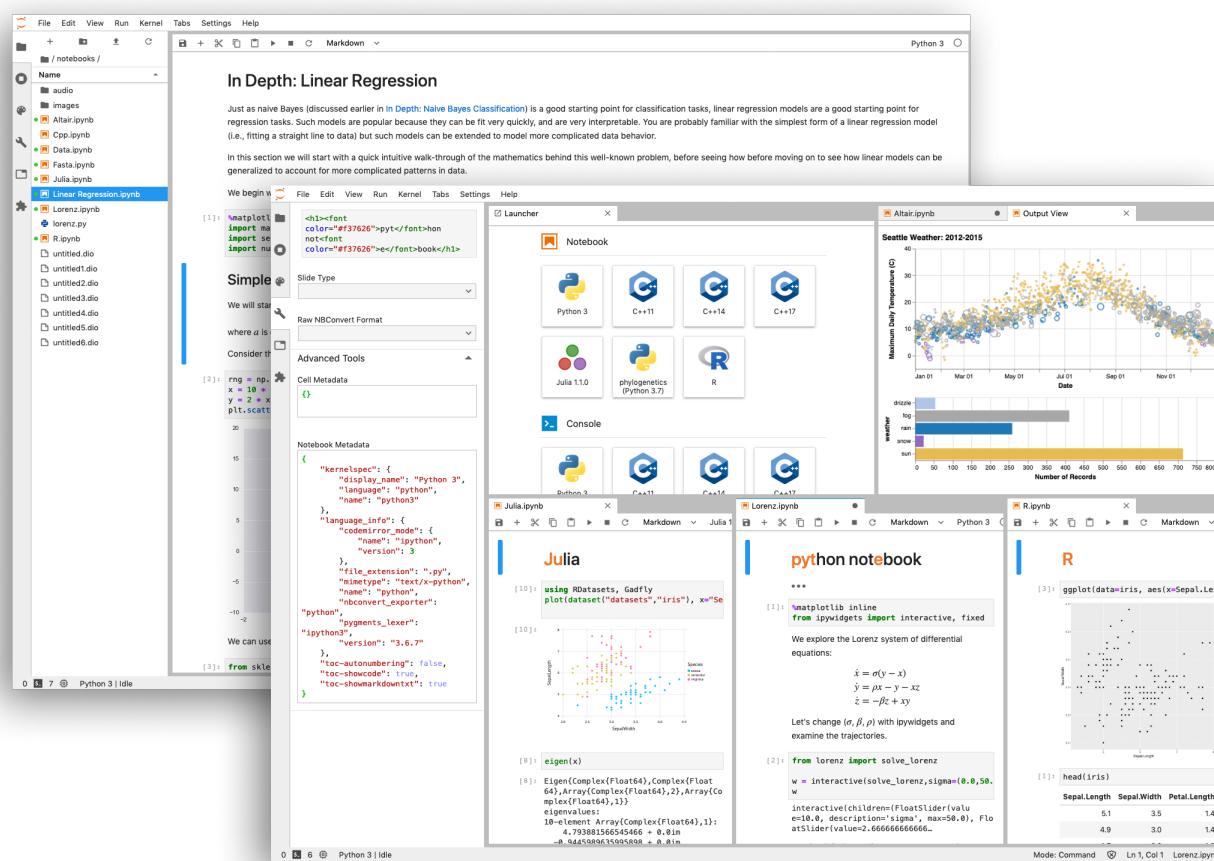
- Anaconda kann eine Diva sein (Häufig Snapshots fahren).
- HDBCLI Bestandteil der HANA Client Tools. Nicht offiziell separat erhältlich.  
HXEDownloadManager notwendig.
- Lösung: Mein Installer script verwenden.  
[https://s3.amazonaws.com/sap\\_s3/hana\\_ml-1.0.7.tar.gz](https://s3.amazonaws.com/sap_s3/hana_ml-1.0.7.tar.gz)

# Jupyter Notebook

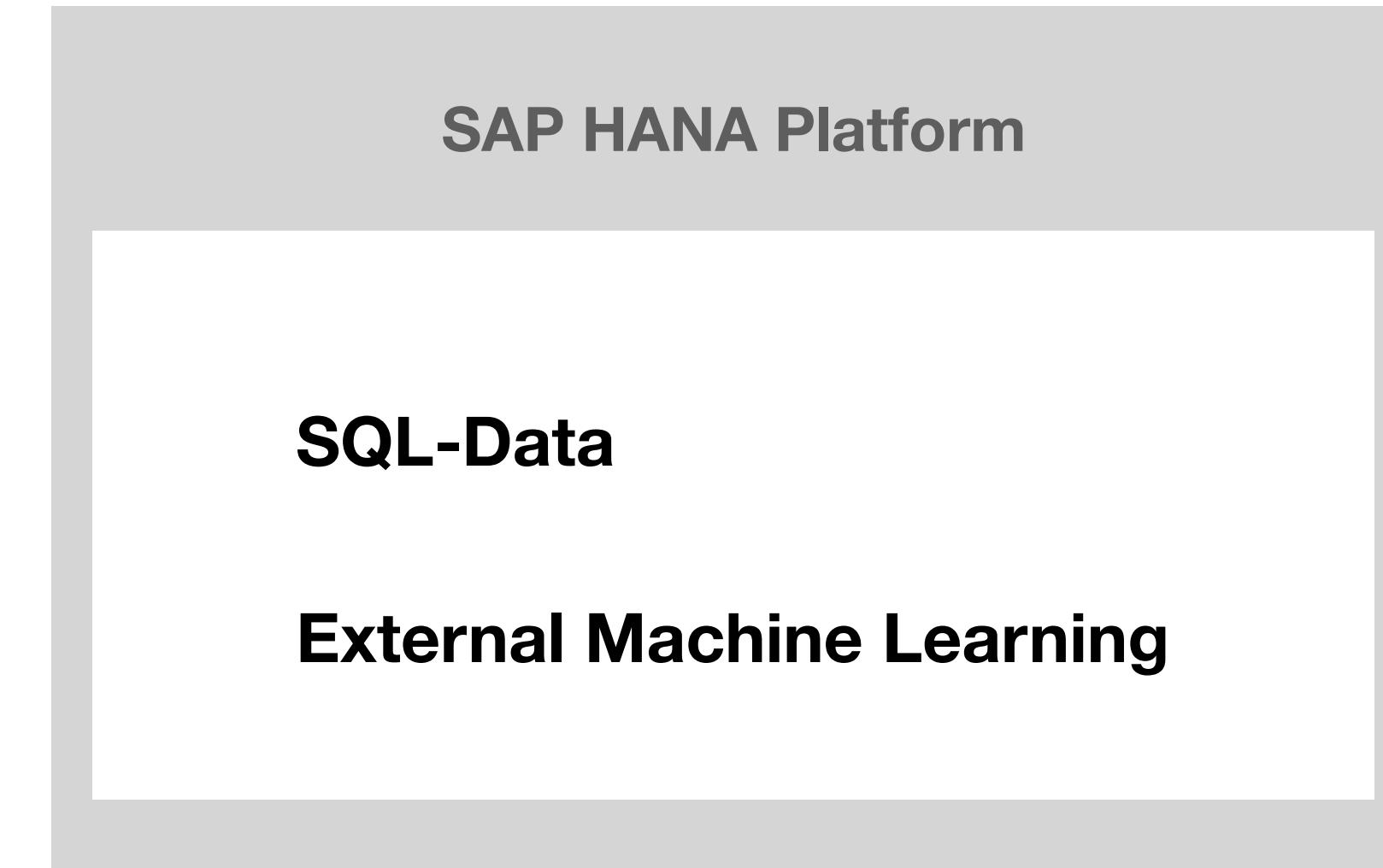
Demo

# Erweiterungsmöglichkeiten der SAP HANA

SAP HANA bietet unterschiedliche Möglichkeiten um externe Tools zu Integrieren



ODBC  
Python



EML call TensorFlow

# Überblick

- **101 Machine Learning**
- **Erweiterungsmöglichkeiten der SAP HANA**
- **Jupyter Notebook**
- **TensorFlow**
- **Beispiel**

# Tensorflow

OpenSource Framework von Google. Sehr populär in Verbindung mit Machine Learning.  
Wird von Google unteranderem in Google Search, Spracherkennung, Maps verwendet.

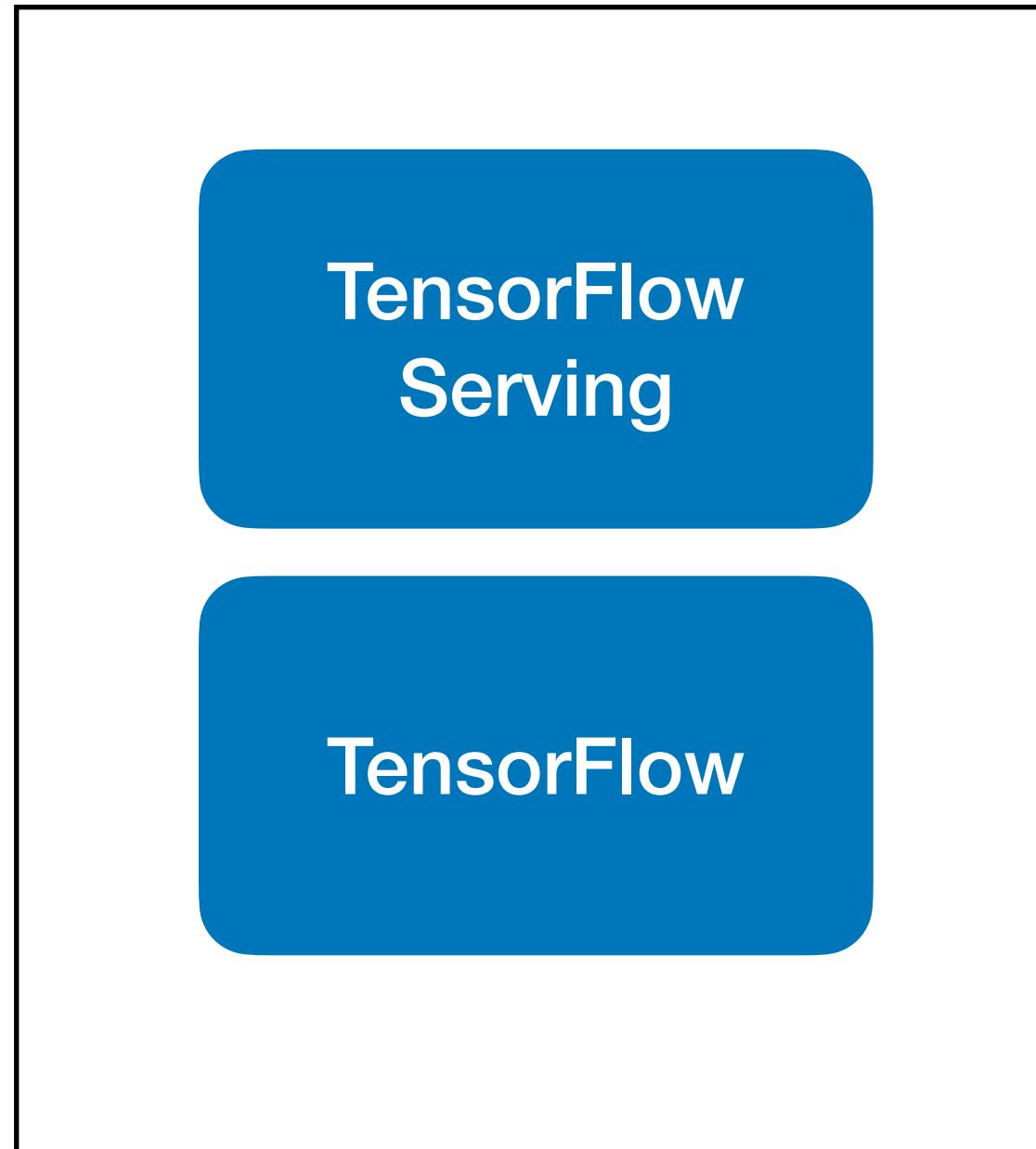


## Key Facts:

- Python interface
- Keras (Deep-Learning Library) seit 1.4
- GPU Beschleunigung
- Tensorflow lite → Nur ausführen von Modellen kein Learning
- Tensorflow Model Serving

# Tensorflow

- Installation

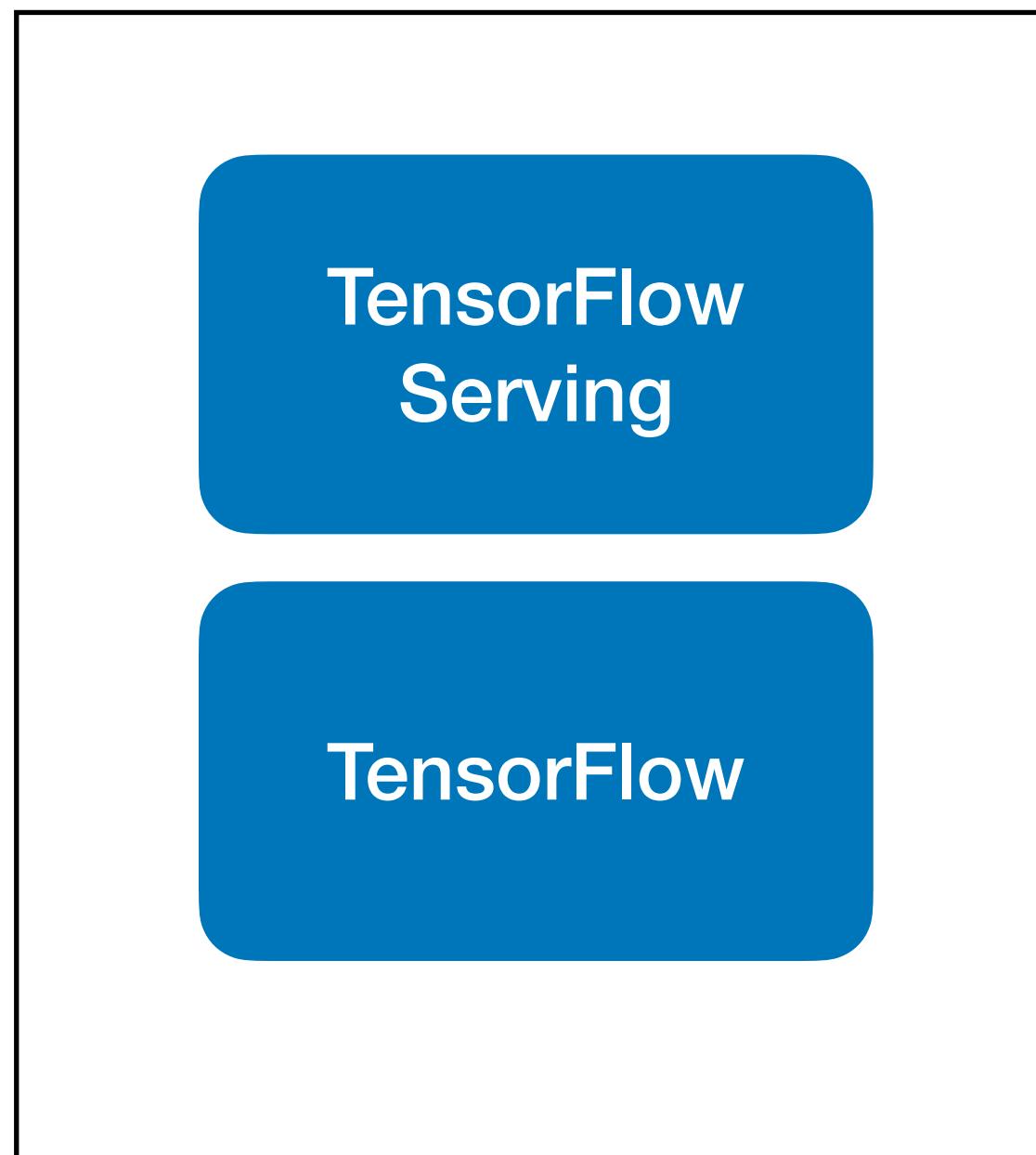


## Schritte

- Linux Container (Ubuntu 18.04)
- Tensorflow installieren. Am besten wieder Anaconda verwenden.
- Tensorflow Serving herunterladen und installieren

# Tensorflow

- Installation



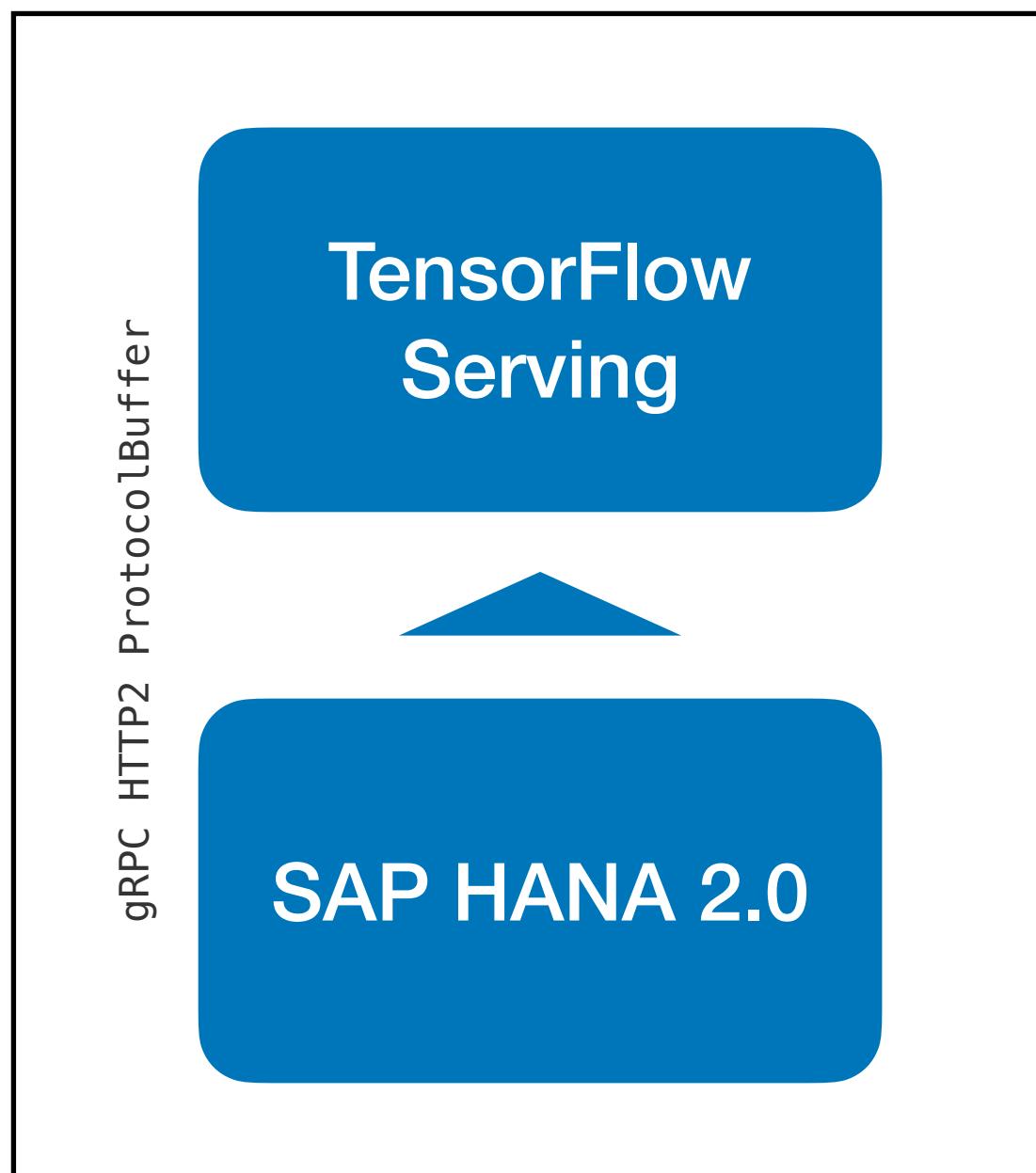
*Container*

## Probleme:

- HANA VM unterstützt nur SUSE und Redhat. Tensorflow unterstützt kein SUSE und Redhat. → Eigene VM für TensorFlow.
- Versionsprobleme mit TensorFlow 2.0 und Serving (Können anscheinend ignoriert werden).

# HANA für Tensorflow vorbereiten

Auch auf HANA Seite muss auch noch etwas getan werden



## Schritte:

- SAP HANA External Machine Learning Library installieren.
- Am besten eigenen User einrichten “ML\_USER”.
- Verbindung zu TensorFlow Serving einrichten:

SQL

```
1 | CREATE REMOTE SOURCE "TensorFlow" ADAPTER "grpc" CONFIGURATION 'server=<ECS Container IP address>;po  
rt=8500';
```

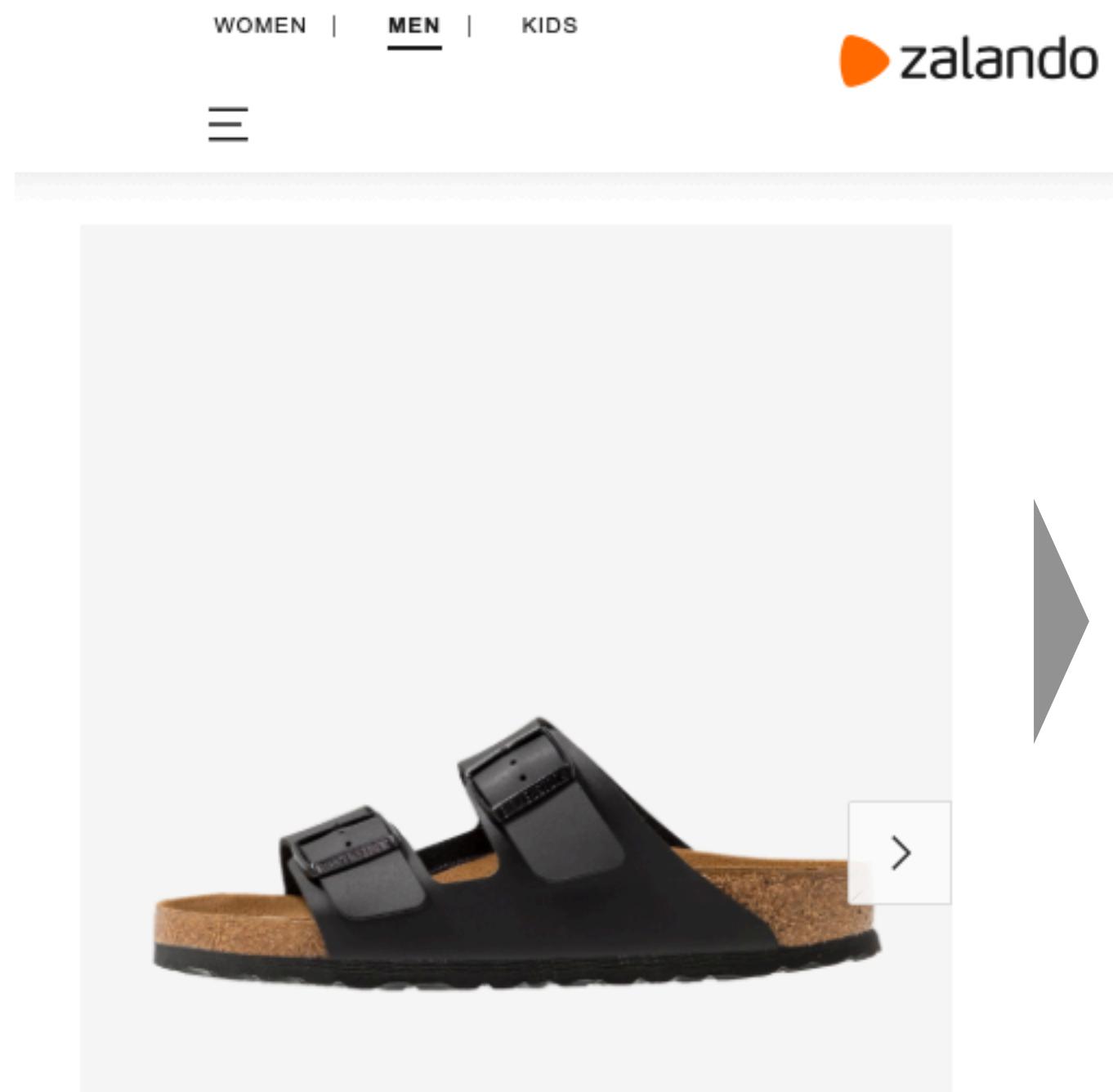
Copy

# Überblick

- **101 Machine Learning**
- **Erweiterungsmöglichkeiten der SAP HANA**
- **Jupyter Notebook**
- **TensorFlow**
- **Beispiel**

# Beispiel

# Fashion Bilderkennung



## Wir wollen:

- Unser System soll Kleidungsstücke automatisch kategorisieren können indem wir ihm ein Foto des Produktes geben

## Lernen:

- Hierfür füttern wir unser Model mit vorbereiten Bildern welche Labels haben
- Datengrundlage:  
Fashion-MINST. 60,000 Bilder von Zalando. (ML-Benchmark)

# Fashion-MNIST



## Vorarbeit

- Die Bilder werden in Graustufen umgewandelt und auf 25x25 verkleinert
- Weniger Speicher, schnellere Verarbeitung

# Jupyter Notebook

Demo

# Fazit

- **Wie aufwendig ist es eine Eigene Machine Learning Plattform aufzusetzen**
  - > Eigener Server mit viel Memory, Virtualisierung für SAP-HANA, Virtualisierung für ML-Plattform
  - > Eigener Proxy, Router
- **Wie kann ich SAP HANA mit externen Tools verbinden und die Plattform so umfassender einfacher verwenden**
  - > Relativ einfach z.B. mit Jupyter Notebook
- **Kann ich die ML-Fähigkeiten von SAP HANA selber erweitern und wie aufwendig ist dies**
  - > Ja Dank EML-API und z.B. Model Serving

# Wer bin ich?

**Twitter:** [@alexander\\_ramm](https://twitter.com/alexander_ramm)

**SAP:** [people.sap.com/pureconsulting\\_alexander.ramm](https://people.sap.com/pureconsulting_alexander.ramm)

## Links:

[Install SAP HANA 2.0, express edition on a Preconfigured Virtual Machine](#)

[Install the Optional SAP HANA External Machine Learning Library Package for SAP HANA, express edition \(Preconfigured VM\)](#)

