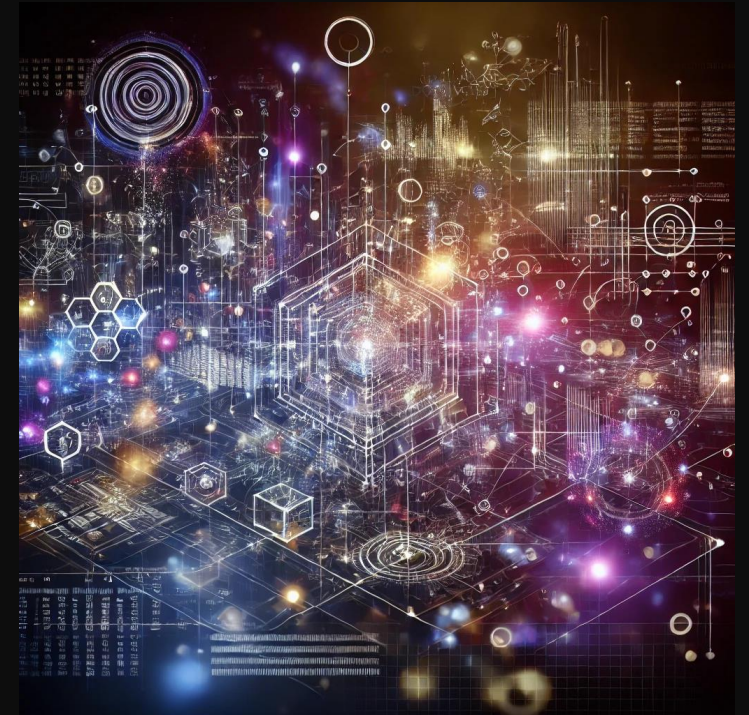


# Data Mining Spring 2025

*Academic Lecturer: Dr. A. Schelle,  
Constructor University*



**C>ONSTRUCTOR**  
UNIVERSITY



# Overview

Lecture 1:  
Introduction to Data  
Mining

Lecture 2: Big Data  
Processes

Lecture 3:  
Statistical  
Processes

Lecture 4: Big Data  
Methods and  
Technologies

Lecture 5: Basic IT  
Law

Lecture 6:  
Principles  
of Machine Learning

Lecture 7: Some  
Elements of Logic

Lecture 8: Uniform  
Search Principle

Lecture 9: Neuronal  
Network Models

Lecture 10: Data  
Mining und ML

Lecture 11:  
Development of AI  
Methods

Lecture 12:  
Selected AI  
Methods

Lecture 13:  
Elementary  
Quantum Field  
Models

# Data Mining in Sprint 2025

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*



Lecture 1



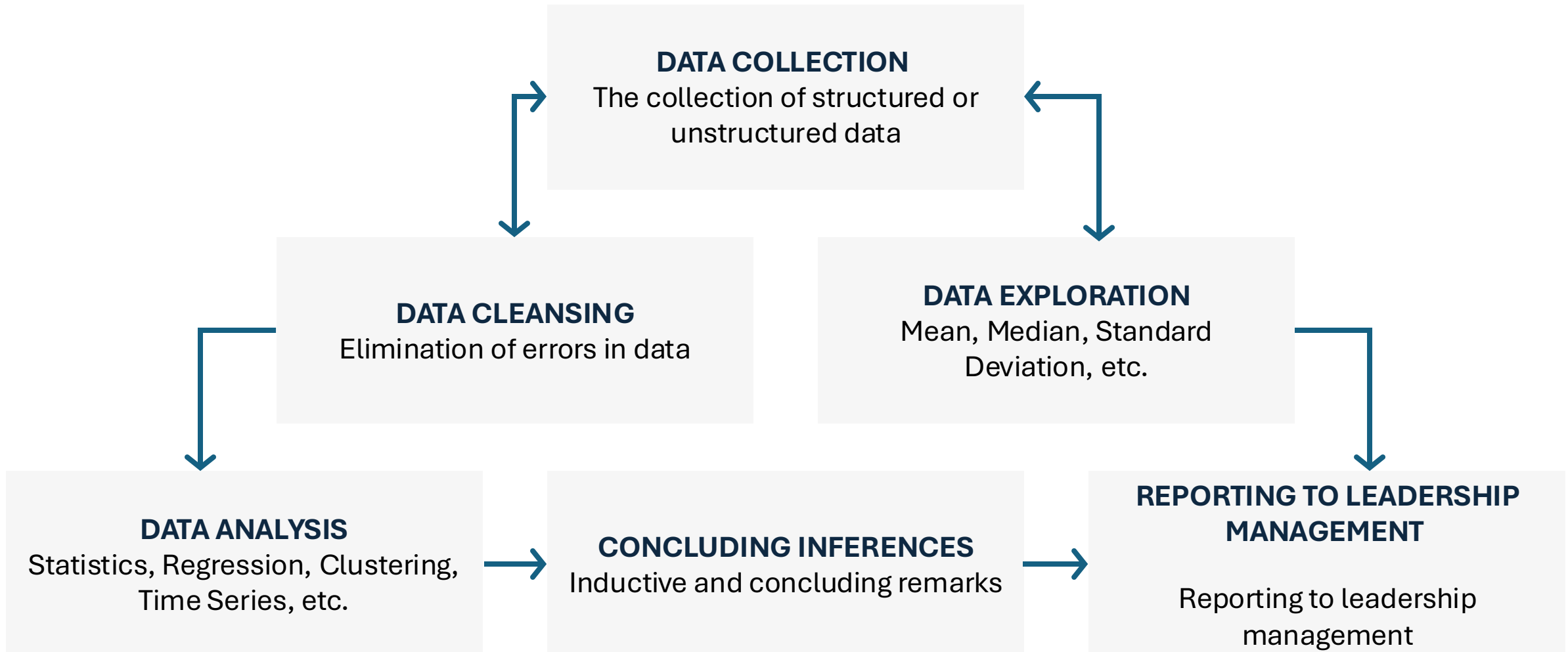
Data Mining



Introduction to Data Mining

# Data Mining Spring 2025

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

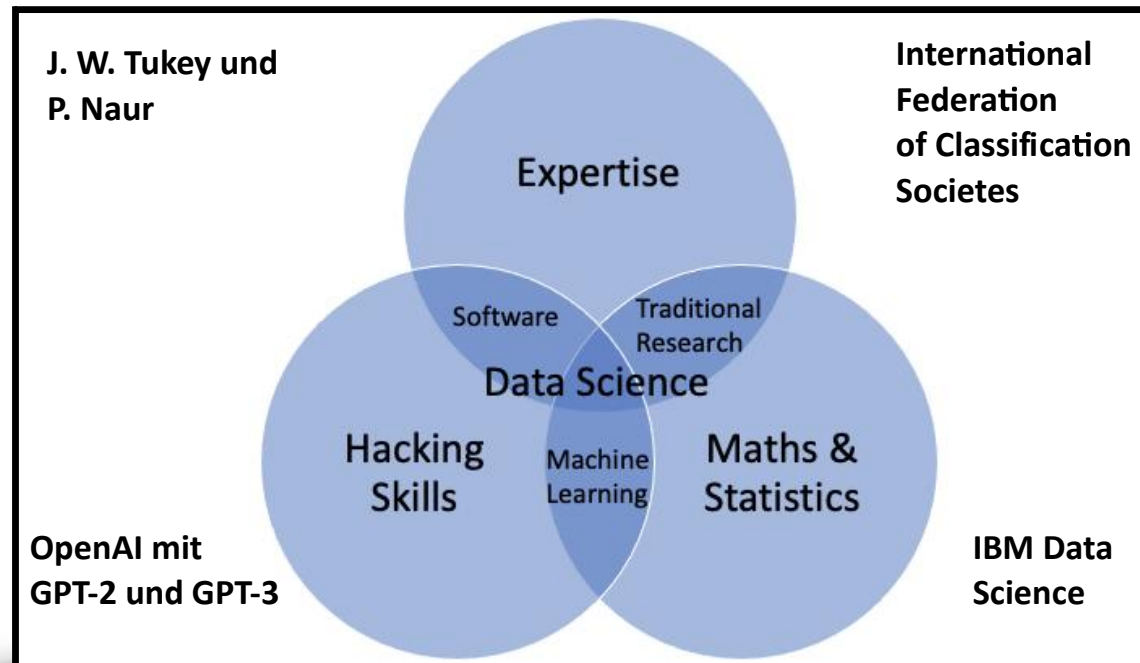


# Data Mining Spring 2025

Academic Lecturer: Dr. A. Schelle, Constructor University

DATA SCIENCE VENN DIAGRAMM

Venn Diagramm: Drew Conway



Quelle : IU Internationale Hochschule (Schelle, 2023)

# HISTORICAL DEVELOPMENT OF DATA SCIENCE AND AI

## Zeitraahmen

<b>Manual Data Processing</b>	10th century
<b>Book Printing</b>	15th century
<b>Index Cards</b>	17th century
<b>Punch Cards</b>	19th century
<b>Mechanical Calculating Machines</b>	19th century
<b>Electronic Data Processing</b>	20th century
<b>Databases</b>	1960s
<b>Data Warehousing</b>	1970s

# HISTORICAL DEVELOPMENT OF DATA SCIENCE AND AI

Zeitraahmen	
Data Mining	1990s
Cloud Computing	2000s
Künstliche Intelligenz & Machine Learning	2010s
Internet of Things	2020s

# Data Science Life Cycle

*Academic Lecturer: Dr. A. Schelle, Constructor University*

Find Data

Prepare Data

Model Data

Model Training

Test Data

Analyse Data





# Data Science Life Cycle – Find Data

*Academic Lecturer: Dr. A. Schelle, Constructor University*

Kaggle.com

Web Scraping

World Bank Data

U.S. Government Data

APIs

***Examples include:***

- **Twitter API:** For social media data analysis.
- **Google Maps API:** For location-based data.
- **OpenWeather API:** For weather-related data.

Compare also Kaggle.com or Github.com.

# Data Science Life Cycle – Prepare Data

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

Data Cleaning

Data  
Transformation

Data Splitting

Data Integration

Feature Engineering

Compare W3Schools.com.



# Data Science Life Cycle – Model Data

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

Linear Regression



Clustering

Decision Tree

Compare Sklearn at  
[ScikitLearn.com](https://scikit-learn.org/).

Neuronal Networks

# Data Science Life Cycle – Model Training

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

Supervised  
Learning

Gradient Descent

Ensemble  
Learning

Cross-Validation

Transfer Learning

Compare Sklearn at  
[ScikitLearn.com](https://scikit-learn.org/).

# Data Science Life Cycle – Test Data

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

Train-Test Split

Confusion Matrix

Performance  
Metrics

Classification

Cross-Validation

ROC Curve and  
AUC

For Regression

Compare W3Schools.com.

# Data Science Life Cycle – Analyse Data

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

Model Evaluation  
Metrics

Feature  
Importance

Cross Validation  
Analysis

Residual Analysis

Learning Curves

Compare Sklearn at [ScikitLearn.com](https://scikit-learn.com).

# 01

**DATA**

Structured And  
Unstructured Data.  
Data Lakes.

# 02

**INFORMATION**

Extraction of  
Knowledge from  
Data Mining.

# 03

**DATA MINING**

Exploration of  
Data Daten  
from Data Mining.

# 04

**DECISIONS**

Influence of work  
and processes  
through information  
from data analysis.

# 05

**OPINIONS**

Subsequent opinions  
after decision and  
presentation of  
information.



# Structured Data - Metadata

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*



A simple example of metadata in English could be:

*For a photo:*

- **Filename:** sunset.jpg
- **Date Taken:** 2025-01-01
- **Resolution:** 1920x1080
- **File Size:** 2.5 MB
- **Camera Model:** Nikon D3500
- **GPS Location:** 34.0522° N, 118.2437° W

```
{ "filename": "sunset.jpg", "dateTaken": "2025-01-01", "resolution": {  
  "width": 1920, "height": 1080 }, "fileSizeMB": 2.5, "cameraModel":  
  "Nikon D3500", "gpsLocation": { "latitude": 34.0522, "longitude": -  
    118.2437 } }
```



# Structured Data – RDFa Daten

*Academic Lecturer: Dr. A. Schelle, Constructor University*



```
html
<!DOCTYPE html>
<html lang="en" xmlns="http://www.w3.org/1999/xhtml" xmlns:dc="http://purl.org/dc/elements/1.1/">
<head>
  <title>RDFa Example</title>
</head>
<body>
  <div vocab="http://schema.org/" typeof="Book">
    <h1 property="name">To Kill a Mockingbird</h1>
    <p>
      Author: <span property="author">Harper Lee</span>
    </p>
    <p>
      Published: <span property="datePublished">1960</span>
    </p>
    <p>
      ISBN: <span property="isbn">978-0061120084</span>
    </p>
    
    <p>
      Description: <span property="description">A novel about racial injustice in the South</span>
    </p>
  </div>
</body>
</html>
```

# Compare Schema.org.

# Structured Data - Example CSV-Data

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

## Example Data in Excel Format (with Image)

### 1. Data Table:

ID	Name	Age	City	Email	Image
1	John Doe	28	New York	<a href="mailto:johndoe@example.com">johndoe@example.com</a>	
2	Jane Smith	34	Los Angeles	<a href="mailto:janesmith@example.com">janesmith@example.com</a>	

Compare W3Schools.com.

# Structured Data - Microdata with Image and HTML

Academic Lecturer: Dr. A. Schelle, Constructor  
University

## Example of Microdata with Image

```
html
Copy code

<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Microdata Example</title>
</head>
<body>
  <div itemscope itemtype="http://schema.org/ImageObject">
    <h1 itemprop="name">Sunset Over the Mountains</h1>
    
    <meta itemprop="description" content="A breathtaking view of the sun setting behind mountains." />
    <meta itemprop="creator" content="John Doe">
    <meta itemprop="dateCreated" content="2024-07-15">
    <meta itemprop="width" content="1920">
    <meta itemprop="height" content="1080">
    <meta itemprop="fileFormat" content="image/png">
  </div>
</body>
</html>
```



## Beispiel für Microdata im HTML:

```
html
Copy code

<!DOCTYPE html>
<html lang="de">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Beispiel für Microdata</title>
</head>
<body>
  <div itemscope itemtype="https://schema.org/Product">
    <h1 itemprop="name">Schöne Sonnenbrille</h1>
    
    <p itemprop="description">Eine stilvolle Sonnenbrille, die Ihren Sommerlook vervollständigt.</p>
    <span itemprop="brand" itemscope itemtype="https://schema.org/Brand">
      Marke: <span itemprop="name">LuxSun</span>
    </span>
    <p>Preis: <span itemprop="price" content="49.99">49,99 €</span></p>
    <meta itemprop="sku" content="12345" />
    <p>Verfügbarkeit: <span itemprop="availability" content="InStock">Auf Lager</span></p>
  </div>
</body>
</html>
```

Compare W3Schools.com.

# Structured Data - JSON Daten

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

## Beispiel für JSON-Daten:

json

 Copy code

```
{
  "name": "Max Mustermann",
  "alter": 29,
  "adresse": {
    "straße": "Musterstraße 1",
    "stadt": "Musterstadt",
    "postleitzahl": "12345"
  },
  "telefonnummern": [
    "+49 123 456789",
    "+49 987 654321"
  ],
  "hobbys": ["Fußball", "Lesen", "Reisen"]
}
```



Compare W3Schools.com.



# Structured Data - Python Dictionary

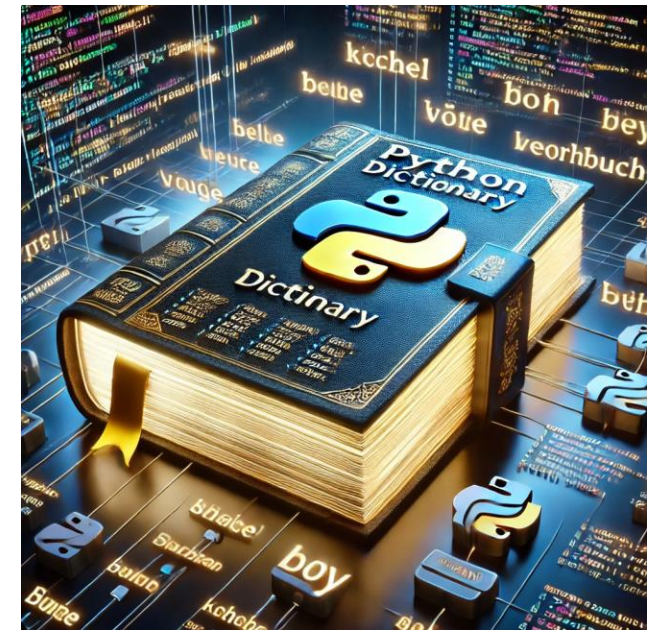
*Academic Lecturer: Dr. A. Schelle, Constructor University*

Beispiel für ein Python Dictionary:

python

 Copy code

```
person = {
    "name": "Anna Müller",
    "alter": 30,
    "beruf": "Webentwicklerin",
    "adresse": {
        "straße": "Beispielstraße 12",
        "stadt": "Berlin",
        "postleitzahl": "10115"
    },
    "telefonnummern": [
        "+49 170 1234567",
        "+49 30 9876543"
    ],
    "verheiratet": False,
    "hobbys": ["Joggen", "Lesen", "Fotografie"]
}
```



# Unstructured Data

*Academic Lecturer: Dr. A. Schelle, Constructor University*



# Big Data, IoT and Data Mining

## BIG DATA

Large Data Volume.

## INTERNET OF THINGS

Collect, Exchange and Edit Data.

## DATA MINING

Extraction of Knowledge from  
Data.

- Volume
- Diversity
- Velocity
- Database Technologies:  
Hadoop,  
NoSQL\_Datenbanken,  
Spark

- Data Stream
- Machine Current
- Location-based Data
- Health Monitoring
- Environmental Monitoring

- Statistics
- Machine Learning
- Artificial Intelligence
- Data Mining
- Time Series Analysis

# Performance Records for Big Data

Improvement of  
Forecasts

Enhancement of  
Visitors

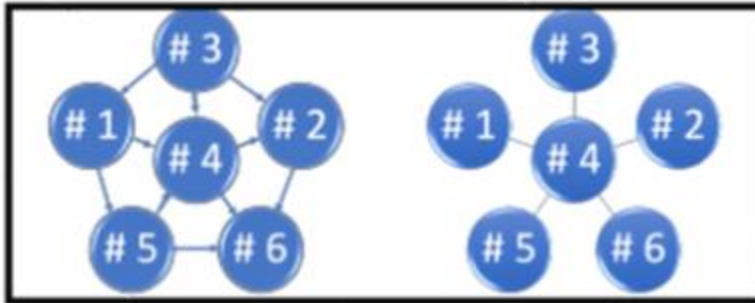
ROC Diagrams

Value Increase

Faster Model  
Building



Darstellung eines gerichteten (links) und eines ungerichteten (rechts) Graphen



Quelle : IU Internationale Hochschule, S. 62 (Schelle, Vogler, Pumperla, 2022).

$$U = (E, K, h)$$

**Matrixdarstellung von  $f$  ist nicht symmetrisch**

$$G = (E, K, f)$$

**Matrixdarstellung von  $f$  ist symmetrisch**

Beispiel eines Referenzengraphen aus arXiv.org.

## Connected Papers

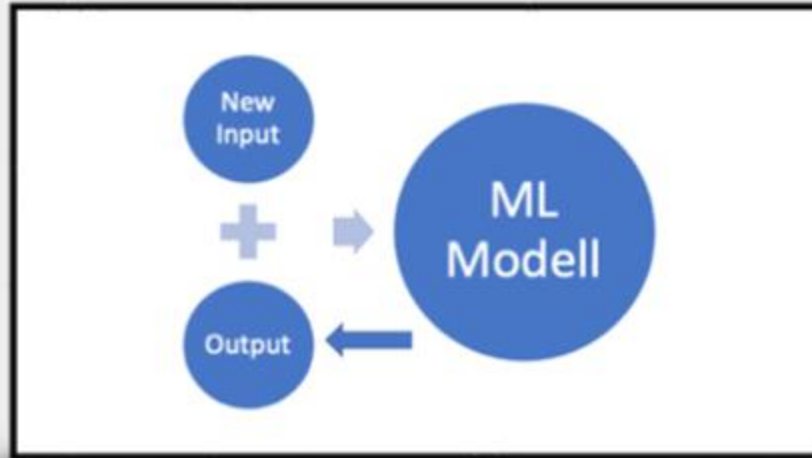


See related papers to:

Microwave-driven atoms: from Anderson localization to Einstein's photoeffect.

Quelle : arXiv-Server. <https://arxiv.org/abs/0809.1082> (2023).

# FUNCTIONALITY OF ARTIFICIAL INTELLIGENCE



Quelle : IU Internationale Hochschule (Schelle, 2023)



**Python, R und SQL**

**Datenstrukturen in  
JSON, JSON-LD,  
RDFa und XML**

**Audio, Bild, Texte  
und Videodateien**

**Gemischte Data  
Lakes**

# OPEN SOURCE SOFTWARE LICENSES

**MIT License**

**GNU License**

**Apache License**

**BSD**

Permits use, modification, distribution,  
with no warranties or liabilities attached.

Allows use, modification, distribution, with conditions and patent grant protections.

Requires modifications to be open-source;  
ensures software freedom and sharing.

Permits use, modification, distribution;  
minimal restrictions, with attribution required.

# Praxisbeispiel 1

Monat	Umsatz (in EUR)	Anzahl der Verkäufe	Durchschnittlicher Warenkorb (in EUR)
Januar	25000	1200	20.83
Februar	28000	1300	21.54
März	32000	1500	21.33
April	29000	1400	20.71
Mai	33000	1600	20.63
Juni	36000	1700	21.18
Juli	38000	1800	21.11
August	40000	1900	21.05
September	35000	1600	21.88
Oktober	37000	1800	20.56
November	41000	2000	20.50
Dezember	43000	2200	20.45

**Tabelle 1:** Tabellarisches Beispiel für die Anzahl der Verkäufe in einem Onlineshop

## Praxisbeispiel: Analyse von Monatlichen Verkaufsdaten

Ein Einzelhandelsunternehmen namens "SuperMart" möchte seine Verkaufsdaten analysieren, um Einblicke in die monatlichen Verkaufstrends zu gewinnen. Das Unternehmen hat Daten über den Umsatz, die verkauften Produkte und die Verkaufszahlen für jeden Monat des letzten Jahres gesammelt.



## Umsatzentwicklung

Die Tabelle zeigt den monatlichen Umsatzverlauf. Die höchsten Umsätze wurden in den Monaten Dezember und November verzeichnet, was auf das Weihnachtsgeschäft und mögliche Sonderangebote hinweisen könnte.

## Verkaufszahlen

Die Anzahl der Verkäufe stieg von Januar bis Dezember kontinuierlich an. Dies könnte auf eine gesteigerte Kundennachfrage, saisonale Veränderungen oder Marketingkampagnen hinweisen.

Der durchschnittliche Warenkorb blieb im Großen und Ganzen konstant, jedoch gab es leichte Schwankungen. Dies könnte auf unterschiedliche Produktmixe in den verschiedenen Monaten zurückzuführen sein.

## Erkenntnisse und Maßnahmen

Angeichts der höchsten Umsätze in den Monaten November und Dezember sollte "SuperMart" sich auf das Weihnachtsgeschäft vorbereiten, indem sie Sonderangebote, Werbeaktionen und Rabatte anbieten.

## Saisonale Schwankungen nutzen

Die steigende Anzahl der Verkäufe von Januar bis Dezember könnte auf saisonale Trends hinweisen. Das Unternehmen könnte seine Inventar- und Marketingstrategien entsprechend anpassen, um von diesen Trends zu profitieren.

## Warenkorb-Optimierung

Obwohl der durchschnittliche Warenkorb relativ stabil ist, könnten gezielte Werbemaßnahmen oder Produktbündelung dazu beitragen, den durchschnittlichen Warenkorbwert weiter zu erhöhen.

## Marketingstrategien überprüfen

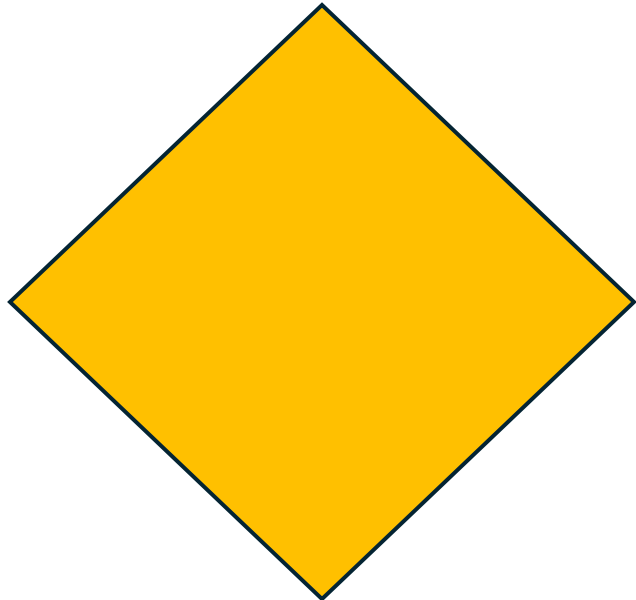
Die Daten könnten auch genutzt werden, um die Wirksamkeit von Marketingkampagnen in verschiedenen Monaten zu bewerten. Falls bestimmte Monate auffällige Umsatzsteigerungen aufweisen, könnten die zugrundeliegenden Marketingstrategien genauer analysiert und auf andere Monate übertragen werden.

In diesem Beispiel zeigt die Tabellendarstellung der Verkaufsdaten, wie die Analyse von Daten dem Unternehmen dabei hilft, fundierte Geschäftsentscheidungen zu treffen und seine Leistung zu optimieren.

# Data Mining in Spring 2025

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

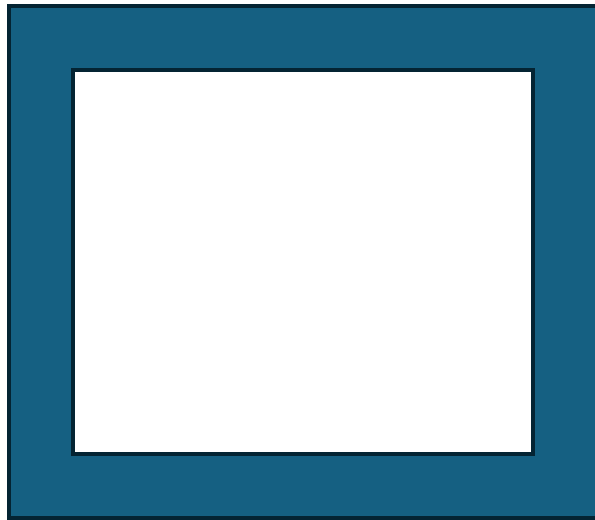
**C>ONSTRUCTOR**  
UNIVERSITY



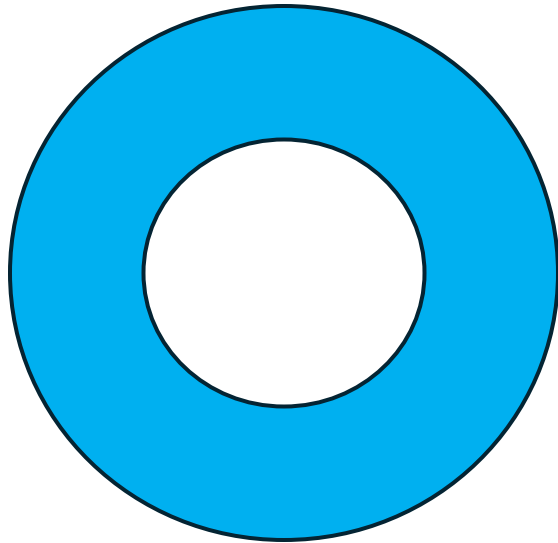
## Summary: Lecture 1

# Data Mining in Spring 2025

*Academic Lecturer: Dr. A. Schelle, Constructor  
University*

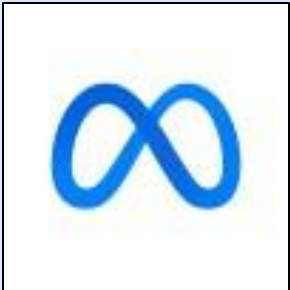


Use Case I: Primary Data  
Mining of Energy  
Consumption Data.



## Task to Lecture 1:

Find a suitable Data Card to develop a Data Mining Model in Python.



# Acknowledgements

*See also LinkedIn.com*

**C>ONSTRUCTOR**  
UNIVERSITY