Data Mining I

Introduction and Course Organisation
Hallo

- **Prof. Dr. Christian Bizer**
- **Professor for Information Systems V**
- **Research Interests:**
  - Data and Web Mining
  - Web Data Integration
  - Data Web Technologies
- **Room:** B6 - B1.15
- **eMail:** chris@informatik.uni-mannheim.de
- **Consultation:** Wednesday, 13:30-14:30

- I will teach the lecture introducing the principle methods of data mining.
Hallo

- M. Sc. Wi-Inf. Anna Primpeli
- Graduate Research Associate
- Research Interests:
  - Semantic Annotations in Web Pages
  - Product Data Integration
  - Identity Resolution
- Room: B6, 26, C 1.04
- eMail: anna@informatik.uni-mannheim.de

Anna will teach exercise 1 (RapidMiner) and will supervise the student projects.
Hallo

- Oliver Lehmberg
- Graduate Research Associate
- Research Interests:
  - Data and Web Mining
  - Network Analysis
  - Web Data Integration
- Room: B6, 26, C 1.04
- eMail: oli@informatik.uni-mannheim.de
- Oliver will teach exercise 2 and 3 (Python) and will supervise the student projects.
Outline of Today‘s Lecture

1. Introduction to Data Mining (70 minutes)
   1. What is Data Mining?
   2. Methods and Applications
   3. The Data Mining Process

2. Course Organisation (20 minutes)
1. Introduction to Data Mining
The Data Deluge

More and more data is generated:

- Transaction data from e-commerce, banking
- Scientific data from astronomy, physics, biology
- Social network sites
- The public Web, twitter, the blogosphere
- Sensor data from machines
- ERP application logs
“We are Drowning in Data...”

Wikipedia = Reference Size

≈ 5.9 TB of data

Source: The following slides are taken from Aidan Hogan's course on “Massive Data Processing”
“We are Drowning in Data...”

Sloan Digital Sky Survey
- ≈ 200 GB/day
- ≈ 73 TB/year
- ≈ 12 Wikipedias/year

Analyze
• Type of sky object:
  Star or galaxy?
“We are Drowning in Data...”

US Library of Congress
≈ 235 TB archived
≈ 40 Wikipedias

Analyze
• Topic distributions
• Citation networks
• Historic trends*.

Facebook
≈ 12 TB/day added
(as of Mar. 2010)
≈ 2 Wikipedias/day
≈ 782 Wikipedias/year

Analyze
• Current interests and behavior of over one billion people.
“We are Drowning in Data...”

Google
\[\approx 20 \text{ PB/day processed} \quad (\text{Jan. 2010})\]
\[\approx 3,389 \text{ Wikipedias/day} \]
\[\approx 7,300,000 \text{ Wikipedias/year} \]

Analyze
• Browsing behavior and interests of users.
“We are Drowning in Data…”

Analyze
- Current behavior and interests of mankind.

This Is What Happens In An
Internet Minute

2018

60 SECONDS

facebook
- 973,000 Logins
- 18 Million Text Messages
- 4.3 Million Videos Viewed

Google
- 3.7 Million Search Queries
- 266,000 Hours Watched

NETFLIX
- $862,823 Spent Online
- 25,000 GIFs Sent via Messenger

Snapchat
- 2.4 Million Snaps Created
- 38 Million Messages

Twitter
- 481,000 Tweets Sent
- 1.1 Million Swipes

Instagram
- 174,000 Scrolling Instagram

Amazon Echo
- 67 Voice-First Devices Shipped

Twitch
- 936,073 Views
- 187 Million Emails Sent
“We are Drowning in Data...”

NSA
Unknown amount of communication data from all over the world.

Analyze
• Identify suspects and terrorists.
“We are Drowning in Data...”

← Amount of data that is produced.

← Amount of data that can be looked at by humans.
"...but starving for knowledge!"

The valuable knowledge is “hidden” in the raw data.

Data Mining methods are needed in many cases to

• make sense of data.
• take business decisions based on data.

We are interested in the patterns not the data itself.
1.1 What is Data Mining?

- Definitions

Non-trivial extraction of
  - implicit,
  - previously unknown and
  - potentially useful
information from data.

Exploration & analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns.

- What is needed? Methods that
  1. detect patterns and regularities in data
  2. support business decisions based on data patterns
Origins of Data Mining

- Draws ideas from machine learning, statistics, and database systems.
- Traditional techniques may be unsuitable due to
  - large amount of data
  - high dimensionality of data
  - heterogeneous and distributed nature of data
Data Mining Application Fields

- Business
  - Customer relationship management, marketing, fraud detection, manufacturing, telecom, health care, ...

- Science
  - Data mining helps scientists to formulate hypotheses.
  - Astronomy, physics, drug discovery, social sciences, ...

- Web and Social Media
  - Advertising, search engine optimization, spam detection, web site optimization, sentiment analysis, ...

- Government
  - Surveillance, crime detection, finding tax cheaters, …
Big Data and the Cloud

- Today, everybody can mine large amounts of data at low costs in the cloud.
- Technical realization
  - massive parallelization using hundreds or thousands of machines
  - using tools like Spark, TensorFlow, Mahout
- Open Data
  - Hundreds of portals offer thousands of data sets
  - https://data.wu.ac.at/portalwatch/
  - https://toolbox.google.com/datasetsearch
- Conference
  - O‘Reilly STRATA Conference
  - http://strataconf.com/public/content/home
The Hottest Skills That Got People Hired in 2016

LinkedIn analyzed the skills of members who started new jobs or received interest from recruiters in 2016.

### The Top Skills of 2016 on LinkedIn Germany

<table>
<thead>
<tr>
<th>Rank</th>
<th>Skill</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cloud and Distributed Computing</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Statistical Analysis and Data Mining</td>
<td>+2</td>
</tr>
<tr>
<td>3</td>
<td>SEO/SEM Marketing</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Marketing Campaign Management</td>
<td>NR</td>
</tr>
<tr>
<td>5</td>
<td>Data Engineering and Data Warehousing</td>
<td>+4</td>
</tr>
<tr>
<td>6</td>
<td>Database Management and Software</td>
<td>+1</td>
</tr>
<tr>
<td>7</td>
<td>Software QA and User Testing</td>
<td>-2</td>
</tr>
<tr>
<td>8</td>
<td>Retail Store Operations</td>
<td>+2</td>
</tr>
<tr>
<td>9</td>
<td>Electronic and Electrical Engineering</td>
<td>-7</td>
</tr>
<tr>
<td>10</td>
<td>Channel Marketing</td>
<td>-2</td>
</tr>
</tbody>
</table>

* NR (Not recorded in 2015)

Data Mining Methods

- **Descriptive Methods**
  - Goal: Find patterns in the data.
  - Example: *Which products are often bought together?*

- **Predictive Methods**
  - Goal: Predict unknown values of a variable
    - given observations (e.g., from the past)
  - Example: *Will a person click a online advertisement?*
    - given her browsing history

- **Machine Learning Terminology**
  - descriptive = unsupervised
  - predictive = supervised
1. Clustering [Descriptive]
2. Classification [Predictive]
3. Regression [Predictive]
4. Association Rule Discovery [Descriptive]
5. Time Series Prediction [Predictive, Data Mining II]
6. Sequential Pattern Discovery [Descriptive, Data Mining II]
7. Anomaly Detection [Descriptive, Data Mining II]
1.2.1 Clustering: Definition

- Given a set of data points, each having a set of attributes, and a similarity measure among them, find clusters such that
  - data points in one cluster are more similar to one another
  - data points in separate clusters are less similar to one another

- Similarity Measures
  - Euclidean distance if attributes are continuous
  - Other problem-specific similarity measures

- Goals
  - Intracluster distances are minimized
  - Intercluster distances are maximized

- Result
  - A descriptive grouping of data points
Clustering: Application 1

- **Application area:** Market segmentation
- **Goal:** Divide a market into distinct subsets of customers
  - where any subset may be conceived as a marketing target to be reached with a distinct marketing mix

- **Approach:**
  1. Collect information about customers
  2. Find clusters of similar customers
  3. Measure the clustering quality by observing buying patterns of customers in same cluster vs. those from different clusters
Clustering: Application 2

- Application area: Document Clustering
- Goal: Find groups of documents that are similar to each other based on terms appearing in them.
- Approach
  1. Identify frequently occurring terms in each document.
  2. Form a similarity measure based on the frequencies of different terms.
- Application Example: Grouping of articles in Google News
Goal: Previously unseen records should be assigned a class from a given set of classes as accurately as possible.

Approach: Given a collection of records (training set)
- each record contains a set of attributes
- one of the attributes is the class (label) that should be predicted.
- Find a model for the class attribute as a function of the values of other attributes.
Classification: Example

- **Training set:**
  - "tree"
  - "tree"
  - "tree"
  - "not a tree"
  - "not a tree"
  - "not a tree"

- **Learned model:** "Trees are big, green plants without wheels."
Classification: Workflow

Class/Label Attribute

Training Set

<table>
<thead>
<tr>
<th>Tid</th>
<th>Attrib1</th>
<th>Attrib2</th>
<th>Attrib3</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>Large</td>
<td>125K</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Medium</td>
<td>100K</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>Small</td>
<td>70K</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>Medium</td>
<td>120K</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>Large</td>
<td>95K</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>Medium</td>
<td>60K</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Yes</td>
<td>Large</td>
<td>220K</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>No</td>
<td>Small</td>
<td>85K</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>No</td>
<td>Medium</td>
<td>75K</td>
<td>No</td>
</tr>
<tr>
<td>10</td>
<td>No</td>
<td>Small</td>
<td>90K</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Unseen Records

<table>
<thead>
<tr>
<th>Tid</th>
<th>Attrib1</th>
<th>Attrib2</th>
<th>Attrib3</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>No</td>
<td>Small</td>
<td>55K</td>
<td>?</td>
</tr>
<tr>
<td>12</td>
<td>Yes</td>
<td>Medium</td>
<td>80K</td>
<td>?</td>
</tr>
<tr>
<td>13</td>
<td>Yes</td>
<td>Large</td>
<td>110K</td>
<td>?</td>
</tr>
<tr>
<td>14</td>
<td>No</td>
<td>Small</td>
<td>95K</td>
<td>?</td>
</tr>
<tr>
<td>15</td>
<td>No</td>
<td>Large</td>
<td>67K</td>
<td>?</td>
</tr>
</tbody>
</table>
Classification: Application 1

- Application area: Fraud Detection
- Goal: Predict fraudulent cases in credit card transactions.
- Approach:
  1. Use credit card transactions and information about account-holders as attributes.
     - When and where does a customer buy? What does he buy?
     - How often he pays on time? etc.
  2. Label past transactions as fraud or fair transactions. This forms the class attribute.
  3. Learn a model for the class attribute from the transactions.
  4. Use this model to detect fraud by observing credit card transactions on an account.
Classification: Application 2

- Application area: Direct Marketing
- Goal: Reduce cost of mailing by targeting the set of consumers likely to buy a new cell-phone product.
- Approach:
  1. Use the data for a similar product introduced before.
     - We know which customers decided to buy and which decided otherwise.
     - This \{buy, don't buy\} decision forms the class attribute.
  2. Collect various demographic, lifestyle, and company-interaction related information about all such customers.
     - Age, profession, location, income, marriage status, etc.
  3. Use this information as input attributes to learn a classification model.
1.2.3 Regression

- Predict a value of a given continuous valued variable based on the values of other variables, assuming a linear or nonlinear model of dependency.

- Greatly studied in statistics and neural network field.

- Examples:
  - Predicting sales amounts of new product based on advertising expenditure.
  - Predicting wind velocities as a function of temperature, humidity, air pressure, etc.
  - Predicting the realizable price of a house or car.

- Difference to classification: The class attribute is continuous, while classification is used for nominal class attributes (e.g., yes/no).
### 1.2.4 Association Rule Discovery: Definition

- Given a set of records each of which contain some number of items from a given collection
- produce dependency rules which will predict occurrence of an item based on occurrences of other items

<table>
<thead>
<tr>
<th>TID</th>
<th>Items</th>
<th>Rules Discovered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bread, Coke, Milk</td>
<td>{Diaper, Milk} --&gt; {Beer}</td>
</tr>
<tr>
<td>2</td>
<td>Beer, Bread</td>
<td>{Milk} --&gt; {Coke}</td>
</tr>
<tr>
<td>3</td>
<td>Beer, Coke, Diaper, Milk</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Beer, Bread, Diaper, Milk</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Coke, Diaper, Milk</td>
<td></td>
</tr>
</tbody>
</table>
Application area: Supermarket shelf management.
- Goal: To identify items that are bought together by sufficiently many customers.
- Approach: Process the point-of-sale data collected with barcode scanners to find dependencies among items.
- A classic rule and its implications:
  - If a customer buys diapers and milk, then he is likely to buy beer as well.
  - So, don’t be surprised if you find six-packs stacked next to diapers!
  - Promote diapers to boost beer sales.
  - If selling diapers is discontinued, this will affect beer sales as well.

Application area: Sales Promotion

Frequently Bought Together

amazon.com

Price For All Three: $87.41
Add all three to Cart  Add all three to Wish List
Show availability and shipping details
Application area: Advertising

Real example:
- Target (American grocery store)
- Analyzes customer buying behavior
- Sends personalized advertisements and coupons

Famous case in the USA:
- Teenage girl gets advertisement for baby products
- ... and her father is mad

Articles about this case
- http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html
- http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/
Application area: Inventory Management:

Goal: A consumer appliance repair company wants to anticipate the nature of repairs on its consumer products and keep the service vehicles equipped with right parts to reduce on number of visits to consumer households.

Approach: Process the data on tools and parts required in previous repairs at different consumer locations and discover the co-occurrence patterns.
1.2.5 Sequential Pattern Discovery: Definition

- Given a sequence of events (or sets of events)
- Find typical temporal patterns:
  - 1. (A,B)  2. (C)  3. (D,E)
  - 1. (A)  2. (B,C)  3. (D)
  - 1. (C)  2. (A,D)  3. (B)

- Typical pattern: Event C usually takes place before event D.
Sequential Pattern Mining: Applications 1

- Application area: Marketing
- Recurring customers
  - Book store example: (Twilight) (New Moon) → (Eclipse)

- Sequential patterns allow more fine grained suggestions than frequent pattern mining without sequence information

- Example:
  - mobile phone → charger vs. charger → mobile phone
    - are indistinguishable by frequent pattern mining
  - customers will buy a charger after a mobile phone
    - but not the other way around!
Sequential Pattern Mining: Applications 2

- Application area: Web usage mining
- Input
  - Web server logs
- Patterns
  - typical sequences of pages visited
- Goal: Improve structure and navigation of website
Which Methods are Used in Practice?

1.3. The Data Mining Process

Source: Fayyad et al. (1996)
1.3.1 Selection and Exploration

- **Selection**
  - What data is available?
  - What do I know about the provenance of this data?
  - What do I know about the quality of the data?

- **Exploration**
  - Get an initial understanding of the data
  - Calculate basic summarization statistics
  - Visualize the data
  - Identify data problems such as outliers, missing values, duplicate records
1.3.2 Preprocessing and Transformation

- Transform data into a representation that is suitable for the chosen data mining methods
  - number of dimensions
  - scales of attributes (nominal, ordinal, numeric)
  - amount of data (determines hardware requirements)

- Methods
  - integrate data from multiple sources
  - aggregation, sampling
  - dimensionality reduction / feature subset selection
  - attribute transformation / text to term vector / embeddings
  - discretization and binarization

- Good data preparation is key to producing valid and reliable models.
- Data integration and preparation is estimated to take 70-80% of the time and effort of a data mining project!
1.3.3 Data Mining

- **Input:** Preprocessed Data
- **Output:** Model / Patterns

1. Apply data mining method.
2. Evaluate resulting model / patterns.
3. **Iterate**
   - Experiment with different parameter settings
   - Experiment with multiple alternative methods
   - Improve preprocessing and feature generation
   - Combine/ensemble multiple methods
1.3.4 Deployment

- Use model in the business context.
How Do Data Scientists Spend Their Days?

What data scientists spend the most time doing

- **Building training sets:** 3%
- **Cleaning and organizing data:** 60%
- **Collecting data sets:** 19%
- **Mining data for patterns:** 9%
- **Refining algorithms:** 4%
- **Other:** 5%

Literature Reference for this Chapter


Chapter 1: Introduction

Chapter 2: Data
2. Course Organisation

- Lecture
  - introduces the principle methods of data mining
  - discusses how to evaluate generated models
  - presents practical examples of data mining applications from the corporate and Web context

- Three alternative Exercise Groups
  - students experiment with data sets using RapidMiner or Python

- Project Work
  - teams of six students realize a data mining project
  - teams may choose their own data sets and tasks (in addition, I will propose some suitable data sets and tasks)
  - teams write a summary about their project and present the project results

- Grading
  - 60% written exam, 30% project report, 10% presentation of project results
Course Organisation

- Course Webpage
  - provides up-to-date information, lecture slides, and exercise material
  - https://www.uni-mannheim.de/dws/teaching/course-details/courses-for-master-candidates/ie-500-data-mining/

- Solutions to the Exercises
  - ILIAS eLearning System, https://ilias.uni-mannheim.de/

- Time and Location
  - Lecture: Wednesday, 10.15 - 11.45
    Room A5, B144
  - Exercise:
    - Thursday, 10.15 - 11.45
      Room B6, A104 (Anna, RapidMiner)
    - Thursday, 12.00 - 13.30,
      Room B6, A104 (Oliver, Python)
    - Thursday, 13.45 - 15.15,
      Room B6, A104 (Oliver, Python)
# Lecture Contents

| 1. Introduction to Data Mining | What is Data Mining?  
Methods and Applications  
The Data Mining Process |
|-------------------------------|--------------------------------------------------|
| 2. Clustering                 | K-means Clustering, Density-based Clustering,  
Hierarchical Clustering, Proximity Measures |
| 3. Classification            | Nearest Neighbor, Decision Trees,  
Model Evaluation, Rule Learning, Naïve Bayes,  
Neural Networks, Support Vector Machines |
| 4. Regression                | Linear Regression, Nearest Neighbor Regression  
Regression Trees, Time Series |
| 5. Association Analysis      | Frequent Item Set Generation, Rule Generation  
Interestingness Measures |
| 6. Text Mining               | Preprocessing Text, Feature Generation, Feature  
Selection, RapidMiner Text Extension |
| 7. Introduction to Student Projects | Requirements and Organization 
Overview of proposed data sets and tasks |
## Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Wednesday</th>
<th>Thursday</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.02.2019</td>
<td>Introduction to Data Mining</td>
<td>Introduction to RapidMiner/Python</td>
</tr>
<tr>
<td>20.02.2019</td>
<td>Lecture Clustering</td>
<td>Exercise Clustering</td>
</tr>
<tr>
<td>27.02.2019</td>
<td>Lecture Classification 1</td>
<td>Exercise Classification</td>
</tr>
<tr>
<td>06.03.2019</td>
<td>Lecture Classification 2</td>
<td>Exercise Classification</td>
</tr>
<tr>
<td>13.03.2019</td>
<td>Lecture Classification 3</td>
<td>Exercise Classification</td>
</tr>
<tr>
<td>20.03.2019</td>
<td>Lecture Regression</td>
<td>Exercise Regression</td>
</tr>
<tr>
<td>27.03.2019</td>
<td>Lecture Text Mining</td>
<td>Exercise Text Mining</td>
</tr>
<tr>
<td>03.04.2019</td>
<td>Introduction to Student Projects and Group Formation</td>
<td>Preparation of Project Outlines</td>
</tr>
<tr>
<td>10.04.2019</td>
<td>Lecture Association Analysis</td>
<td>Exercise Association Analysis</td>
</tr>
<tr>
<td></td>
<td>- Easter Break -</td>
<td></td>
</tr>
<tr>
<td>01.05.2019</td>
<td>- Holiday -</td>
<td>Feedback on demand</td>
</tr>
<tr>
<td>08.05.2019</td>
<td>Project Work</td>
<td>Feedback on demand</td>
</tr>
<tr>
<td>15.05.2019</td>
<td>Project Work</td>
<td>Feedback on demand</td>
</tr>
<tr>
<td>22.05.2019</td>
<td>Project Work</td>
<td>Submission of project results</td>
</tr>
<tr>
<td>29.05.2019</td>
<td>Presentation of project results</td>
<td>- Holiday -</td>
</tr>
<tr>
<td>03.06.2019</td>
<td>Final Exam</td>
<td></td>
</tr>
</tbody>
</table>
Deadlines

- Submission of project proposal
  - Sunday, April 7\textsuperscript{th}, 23:59

- Submission of final project report
  - Sunday, May 26\textsuperscript{th}, 23:59

- Project presentations
  - Thursday, May 30\textsuperscript{th}
  - everyone has to attend the presentations

- Final written exam
  - Monday, June 3\textsuperscript{rd}
Literature

   - main reference book for the course!
   - we provide scans of important chapters via ILIAS

2. Vijay Kotu, Bala Deshpande: *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. Morgan Kaufmann.
   - covers some theory, but also many practical aspects using RapidMiner

   - explains how to apply the covered methods using Python

4. Website: KDnuggets
   - Overview of tools, online courses, events
   - http://www.kdnuggets.com/
Software

- Powerful data mining suite
- We are using Version 9.0 in the exercise
Gartner 2018 Magic Quadrant for Advanced Analytics Platforms
Literature – Rapidminer

1. Rapidminer – Documentation
   • http://docs.rapidminer.com
   • https://academy.rapidminer.com/catalog
   • https://www.youtube.com/user/RapidIVideos

   • covers theory and practical aspects using RapidMiner

   • Explains along case studies how to use simple and advanced Rapidminer features
We use the Anaconda Python distribution, which includes:

- Pandas, Numpy, Matplotlib
- Scikit-learn

Exercises are provided as Jupyter notebooks

You will need to code!

```python
import pandas as pd
from sklearn.model_selection import StratifiedKFold
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix, accuracy_score

cv = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)
naive_bayes = GaussianNB()
target_prediction = cross_val_predict(naive_bayes,
    credit_data.values, credit_target, cv=cv)

cm = confusion_matrix(credit_target, target_prediction)
cost = cm[0][1] * 100 + cm[1][0] * 1
acc = accuracy_score(credit_target, target_prediction)

print("Naive Bayes with accuracy of {} and cost {}\n".format(acc, cost))
print(confusion_matrix_report(credit_target, target_prediction))
```
Literature – Python

- **Python tutorial**: [https://docs.python.org/3/tutorial/](https://docs.python.org/3/tutorial/)
- Great book: **Hands-on Machine Learning with Scikit-Learn** by Aurélien Géron (freely available online via the University library)
Usage of Python versus R versus RapidMiner

Source: KDnuggets online poll, 2900 votes
Lecture Videos and Screencasts

1. Video recordings of all lectures from FSS 2015
2. Step-by-step introduction to relevant RapidMiner features

http://dws.informatik.uni-mannheim.de/en/teaching/lecture-videos/
Advertisement: Career Fair

20.03.2019 – 16:00-18:00 Uhr
MINT-MARKTPLATZ
Fakultät für Wirtschaftsinformatik & Wirtschaftsmathematik
B6, 30-32, Bauteil E-F (Neubau) im 1.OG

PRAKTIKA
FESTANSTELLUNGEN
VORTRÄGE

Accenture
BCG Gamma
Commerzbank
d-fine GmbH
Essity GmbH

EXA Deutschland GmbH
Inter-Versicherung
KPMG
mayato
MSW & Partner

Porsche
Roche Diagnostics
SAP AG
Scheer Group
Stoccard

Xenium
zeb
Questions?