

# How to read a Paper

ISMLL

Eya Boumaiza

# Outline

How to read a paper

Common paper structure

Finding additional material

# Data Analytics Seminar 1 -Anomaly Detection



## How to read a paper

- ▶ Like novel or newspaper stories, scientific articles needs to be read differently.
- ▶ Since they are not books designed for students sometimes they are not self contained and requires some research to be fully understood.
- ▶ Understand a paper for a researcher means to be able to implement the described algorithm.

# Data Analytics Seminar 1 -Anomaly Detection



## How to read a paper

- ▶ Skim
- ▶ Re-read
- ▶ Analyze
- ▶ Summarize

# Data Analytics Seminar 1 -Anomaly Detection



## Skim

First get the "Big picture" by reading the title, abstract, and introduction carefully: this will tell you the major findings and why they matter.

- ▶ Quickly scan the article without taking notes: focus on headings and subheadings
- ▶ Note the publishing date and conference/journal
- ▶ Note terms and parts you don't understand.  
Only with the bigger picture you will understand how much it is necessary to investigate something.

# Data Analytics Seminar 1 -Anomaly Detection



## Re-read

Read the article again, asking yourself questions such as:

- ▶ What problems is the study trying to solve?
- ▶ Are findings well supported by evidence?
- ▶ Is the study repeatable? (i.e. is the article self contained?)
- ▶ If you do not understand take some time to find a brief explanation of what you are not understanding (one-two sentences).
- ▶ Is the paper innovative?

# Data Analytics Seminar 1 -Anomaly Detection



## Interpret

- ▶ Examine graphs and tables carefully
- ▶ Try to interpret data first before looking at captions
- ▶ When reading the discussion and results look after key issues and new findings
- ▶ Make sure you have distinguished the main points. If not go over the text again.

# Data Analytics Seminar 1 -Anomaly Detection



## Summarize

- ▶ Take notes and underline key points: it improves reading
- ▶ Decide what part of the paper needs to be expanded and how much.



# Data Analytics Seminar 1 -Anomaly Detection



## Common paper structure

- ▶ Abstract
- ▶ Introduction
- ▶ State of the art
- ▶ Algorithms explanation
- ▶ Experiments
- ▶ Conclusions and future work
- ▶ References

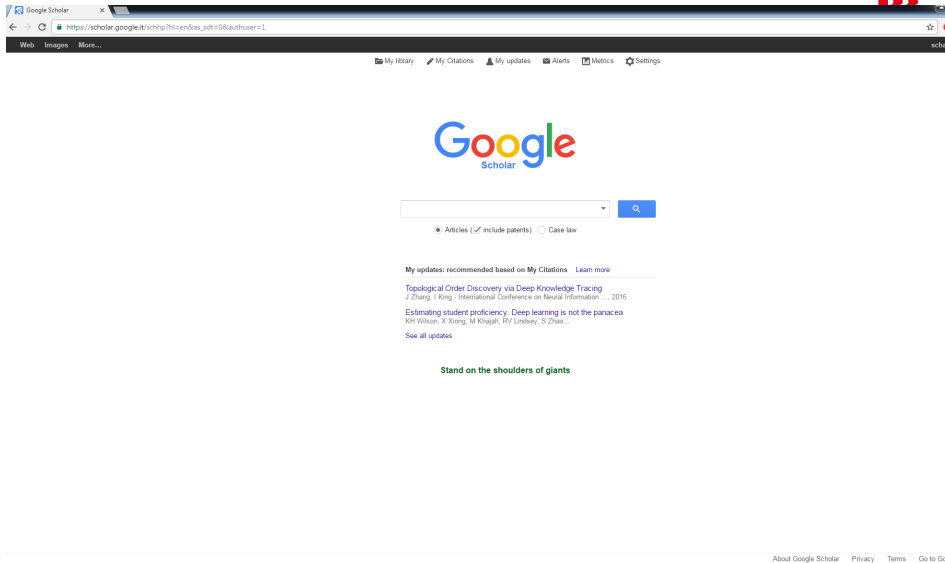
# Data Analytics Seminar 1 -Anomaly Detection



Let's take this paper as an example:

"Huang, S., Wang, S., Liu, T. Y., Ma, J., Chen, Z., and Veijalainen, J. (2015, August). **Listwise Collaborative Filtering**. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 343-352). ACM."

Go to: [www.scholar.google.com/](http://www.scholar.google.com/)



The image shows a screenshot of the Google Scholar homepage. At the top, there is a navigation bar with links for "My library", "My Citations", "My updates", "Alerts", "Metrics", and "Settings". Below this is the Google Scholar logo. A search bar is present with a dropdown menu and a search button. Underneath the search bar, there are radio buttons for "Articles", "include patents" (which is checked), and "Case law". Below the search bar, there is a section for "My updates: recommended based on My Citations" with a "Learn more" link. Two articles are listed: "Topological Order Discovery via Deep Knowledge Tracing" by J Zhang, I King - International Conference on Neural Information ..., 2016, and "Estimating student proficiency: Deep learning is not the panacea" by KH Wilson, X Xiong, M Khajeh, RV Lindsey, S Zhao... Below the articles, there is a link "See all updates". At the bottom of the page, there is a footer with links for "About Google Scholar", "Privacy", "Terms", and "Go to Google Scholar".

Listwise Collaborative Filtering

https://scholar.google.it/scholar?hl=en&authuser=1&q=Listwise+Collaborative+Filtering&btnG=&as\_sdt=1%2C5&as\_sdt=

Web Images More...

Google Listwise Collaborative Filtering

Scholar

About 1,280 results (0.09 sec)

Articles

**Listwise Collaborative Filtering**[\[PDF\] researchgate.net](#)

Case law

[S Huang](#), [S Wang](#), [TY Liu](#), [J Ma](#), [Z Chen](#) - Proceedings of the 38th ..., 2015 - dl.acm.org  
 Abstract Recently, ranking-oriented collaborative filtering (CF) algorithms have achieved great success in recommender systems. They obtained state-of-the-art performances by estimating a preference ranking of items for each user rather than estimating the absolute ...  
 Cited by 2 Related articles All 6 versions Cite Save More

My library

Any time

**List-wise learning to rank with matrix factorization for collaborative filtering**[\[PDF\] tudelft.nl](#)

Since 2016

[Y Shi](#), [M Larson](#), [A Hanjalic](#) - Proceedings of the fourth ACM conference ..., 2010 - dl.acm.org  
 Abstract A ranking approach, ListRank-MF, is proposed for **collaborative filtering** that combines a **list-wise** learning-to-rank algorithm with matrix factorization (MF). A ranked list of items is obtained by minimizing a loss function that represents the uncertainty between ...  
 Cited by 99 Related articles All 16 versions Cite Save More

Since 2015

Since 2012

Custom range...

Sort by relevance

**Learning to rank: from pairwise approach to listwise approach**[\[PDF\] wustl.edu](#)

Sort by date

[Z Cao](#), [T Qin](#), [TY Liu](#), [MF Tsai](#), [H Li](#) - Proceedings of the 24th international ..., 2007 - dl.acm.org  
 ... These include document retrieval, **collaborative filtering**, expert finding, anti web spam, sentiment analysis, and product ... metric between the corresponding top k probability distributions as the **list-wise** loss function ... when we use Cross Entropy as metric, the **listwise** loss function ...  
 Cited by 901 Related articles All 26 versions Cite Save More

 include patents include citations Create alert**Probabilistic latent preference analysis for collaborative filtering**[\[PDF\] cuhk.edu.hk](#)

[NN Liu](#), [M Zhao](#), [Q Yang](#) - Proceedings of the 18th ACM conference on ..., 2009 - dl.acm.org  
 ... To recommend new items to a user, content-based **filters** match their representations to those items the user has expressed interests on. In contrast, the **collaborative filtering**(CF) approach does not require any content information about the items, it works by collecting ratings ...  
 Cited by 78 Related articles All 5 versions Cite Save More

**Effort estimation based on collaborative filtering**[\[PDF\] toyo.ac.jp](#)

[N Ohsugi](#), [M Tsunoda](#), [A Monden](#) - ... Conference on Product ..., 2004 - Springer  
 ... Their results showed that **listwise** deletion technique did not performed well when the level of missing data was more than 30 ... In this paper, we propose **Collaborative Filtering** (CF) based effort estimation method, under the assumption that the (historical) predictor data have a ...  
 Cited by 18 Related articles All 18 versions Web of Science 2 Cite Save More

# Data Analytics Seminar 1 -Anomaly Detection



## Abstract

- ▶ Brief introduction to the topic
- ▶ Brief introduction to paper achievements
- ▶ Brief summary of the experiments

# Data Analytics Seminar 1 -Anomaly Detection



## Introduction

- ▶ Introduction to the topic
- ▶ Introduction of the main concepts
- ▶ Introduction of the main state of the art methods
- ▶ **State of the art limitations**
- ▶ **Hypotheses**
- ▶ **Contributions**

# Data Analytics Seminar 1 -Anomaly Detection



## The Importance of Hypotheses

- ▶ It is not enough to describe some new technique or system, some claim about it must be stated and evaluated
- ▶ In experimental research, hypotheses typically take one of these two forms:
  - ▶ Technique/system X automates task Y for the first time
  - ▶ Technique/system X automates task Y better, along some dimension, than each of its rivals
- ▶ In theoretical papers, the hypotheses are the statements of theorems and the supporting evidence is their proofs

# Data Analytics Seminar 1 -Anomaly Detection



## The Importance of Hypotheses

Technique/system X automates task Y better, along some dimension, than each of its rivals, where the dimensions are typically:

- ▶ **Behavior:** X has a higher success rate or produces better quality outputs than Y
- ▶ **Coverage:** X is applicable to a wider range of examples than Y
- ▶ **Efficiency:** X is faster or uses less space than Y
- ▶ **Dependability:** X is more reliable, safe or secure than its rivals
- ▶ **Maintainability:** X is easier to adapt and extend than its rivals
- ▶ **Usability:** Users find X easier to use than its rivals



# Data Analytics Seminar 1 -Anomaly Detection



## State of the art / Related work

- ▶ Is a broad and shallow account of the field, which helps to place the contribution of the paper in context
- ▶ What are the rival approaches?
- ▶ What are the drawbacks of each?
  - ▶ One sentence per method. Is it clear enough?
- ▶ How has the battle between different approaches progressed?
- ▶ What are the major outstanding problems?

# Data Analytics Seminar 1 -Anomaly Detection



## Algorithm Explanation

- ▶ First the authors introduce the algorithm from which they derived the new algorithm
- ▶ Then, the new algorithm is explained
- ▶ Contains:
  - ▶ Formulas
  - ▶ Pseudo code

---

**Algorithm 1:** The ListCF Algorithm
 

---

**Input:** An item set  $I$ , a user set  $U$ , and a rating matrix  $R \in \mathbb{R}^{M \times N}$ . A set of rated items  $I_u \subseteq I$  by each user  $u \in U$ . The maximal number of iterations  $maxIteration$  and error threshold  $\epsilon$ .

**Output:** A ranking  $\hat{r}_u$  of items for each user  $u \in U$ .

```

1 for  $u \in U$  do
2   for  $v \in U$  and  $u \neq v$  do
3      $P_u, P_v \leftarrow \text{TopKProDist}(I_u, I_v, R)$  /* Eq.1 */
4      $sim(u, v) \leftarrow \text{Similarity}(P_u, P_v)$  /* Eq.2 */
5   end
6    $N_u \leftarrow \text{SelectNeighbors}(\{sim(u, v)\}_{v \in U/u})$ 
7 end
8 for  $u \in U$  do
9    $t = 1$ 
10  repeat
11     $\epsilon = 0$ 
12    Initialize( $\varphi_u^0$ )
13    for  $g \in \mathcal{G}_k^{T_u}$  do
14       $\varphi_{u,g}^t \leftarrow \text{Update}(N_u, sim, R)$  /* Eq.8 */
15       $\epsilon += \sqrt{\sum (\varphi_{u,g}^t - \varphi_{u,g}^{t-1})^2}$ 
16    end
17     $t \leftarrow t + 1$ 
18  until  $t > maxIteration$  or  $\epsilon < \epsilon$ ;
19  for  $t \in T_u$  do
20     $P(t) \leftarrow \text{Aggregation}(\{\varphi_{u,g}\}_{g \in \mathcal{G}_k^{T_u}})$ 
21  end
22   $\hat{r}_u \leftarrow \text{Ordering}(\{P(t)\}_{t \in T_u})$ 
23 end

```

---

# Data Analytics Seminar 1 -Anomaly Detection



## Experiments

### ▶ Dataset Explanation

- ▶ What are the available information?
- ▶ What are the available statistics? E.g. number of users, items, sparsity etc.

### ▶ Evaluation protocol

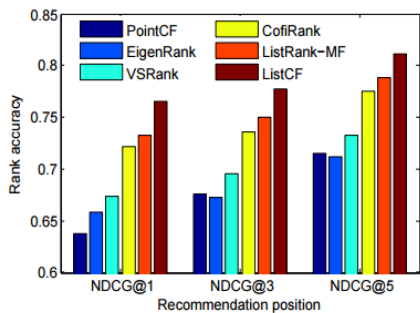
- ▶ How is the error of the algorithm computed?
- ▶ Are there any other quantitative success measures?

### ▶ Experiments

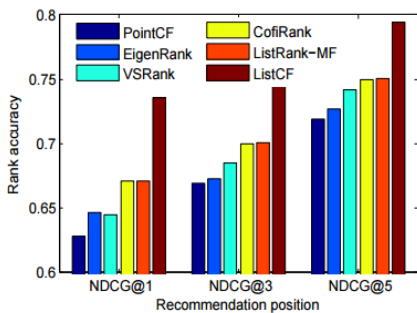
- ▶ Are the results statistically significant?

**Table 2: Statistics on the three datasets.**

	Movielens-1M	EachMovie	Netflix
#users	6,040	36,656	429,584
#items	3,952	1,623	17,770
#ratings	1,000,209	2,580,222	99,884,940
#ratings/user	165.6	70.4	232.5
#ratings/item	253.1	1589.8	5621.0
sparsity	93.7%	95.7%	98.7%



(a) Movielens-1M



(b) EachMovie

# Data Analytics Seminar 1 -Anomaly Detection



## Conclusions

- ▶ Repeats the contributions pointing out specifically how the paper addressed it
- ▶ Include future works

# Data Analytics Seminar 1 -Anomaly Detection



## Finding additional material

- ▶ If you don't understand something..
- ▶ This is not a book, it happens...
  - ▶ Try to pose yourself a specific questions
  - ▶ Look online



# Data Analytics Seminar 1 -Anomaly Detection



## Finding additional material

- ▶ A book explaining the algorithms
- ▶ A PhD thesis
- ▶ Tutorials
- ▶ Highly related state of the art papers

Listwise Collaborative Filtering

https://scholar.google.it/scholar?hl=en&authuser=1&q=Listwise+Collaborative+Filtering&btnG=&as\_sdt=1%2C5&as\_sdtp=

Web Images More...

Google Listwise Collaborative Filtering

Scholar

About 1,280 results (0.09 sec)

Articles

**Listwise Collaborative Filtering**

S. Huang, S. Wang, T.Y. Liu, J. Ma, Z. Chen - Proceedings of the 38th ... 2015 - dl.acm.org

[\[PDF\] researchgate.net](#)

Case law

Abstract Recently, ranking-oriented collaborative filtering (CF) algorithms have achieved great success in recommender systems. They obtained state-of-the-art performances by estimating a preference ranking of items for each user rather than estimating the absolute ...

Cited by 2 Related articles All 6 versions Cite Save More

My library

Any time

**List-wise learning to rank with matrix factorization for collaborative filtering**

Y. Shi, M. Larson, A. Hanjalic - Proceedings of the fourth ACM conference ... 2010 - dl.acm.org

[\[PDF\] tudelft.nl](#)

Since 2016

Abstract A ranking approach, ListRank-MF, is proposed for **collaborative filtering** that combines a **list-wise** learning-to-rank algorithm with matrix factorization (MF). A ranked list of items is obtained by minimizing a loss function that represents the uncertainty between ...

Cited by 99 Related articles All 16 versions Cite Save More

Since 2015

Since 2012

Custom range...

Sort by relevance

**Learning to rank: from pairwise approach to listwise approach**

Z. Cao, T. Qin, T.Y. Liu, M.F. Tsai, H. Li - Proceedings of the 24th international ... 2007 - dl.acm.org

[\[PDF\] wustl.edu](#)

Sort by date

... To recommend document retrieval, **collaborative filtering**, expert finding, anti web spam, sentiment analysis, and product ... metric between the corresponding top k probability distributions as the **list-wise** loss function ... when we use Cross Entropy as metric, the **listwise** loss function ...

Cited by 901 Related articles All 26 versions Cite Save More

 include patents include citations Create alert**Probabilistic latent preference analysis for collaborative filtering**

N.N. Liu, M. Zhao, Q. Yang - Proceedings of the 18th ACM conference on ... 2009 - dl.acm.org

[\[PDF\] cuhk.edu.hk](#)

... To recommend new items to a user, content-based **filters** match their representations to those items the user has expressed interests on. In contrast, the **collaborative filtering**(CF) approach does not require any content information about the items, it works by collecting ratings ...

Cited by 78 Related articles All 5 versions Cite Save More

**Effort estimation based on collaborative filtering**

N. Ohsugi, M. Tsunoda, A. Monden - ... Conference on Product ... 2004 - Springer

[\[PDF\] toyo.ac.jp](#)

... Their results showed that **listwise** deletion technique did not performed well when the level of missing data was more than 30 ... In this paper, we propose **Collaborative Filtering** (CF) based effort estimation method, under the assumption that the (historical) predictor data have a ...

Cited by 18 Related articles All 18 versions Web of Science 2 Cite Save More



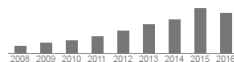
## Sheng-Lung Huang

National Taiwan University  
Biomedical imaging, fiber, laser, crystal  
Verified email at ntu.edu.tw

[Follow](#)
[Google Scholar](#)

[Get my own profile](#)

Citation indices	All	Since 2011
Citations	15313	10955
h-index	49	39
i10-index	408	298



Title	1-20	Cited by	Year
<b>Strain engineering and one-dimensional organization of metal-insulator domains in single-crystal vanadium dioxide beams</b>		266	2009
J Cao, E Ertekin, V Srinivasan, W Fan, S Huang, H Zheng, JWL Yim, ... Nature nanotechnology 4 (11), 732-737			
<b>Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts.</b>		248	1994
P Savatier, S Huang, I Szekeley, KG Wiman, J Samarut Oncogene 9 (3), 809-818			
<b>Flooding-induced membrane damage, lipid oxidation and activated oxygen generation in corn leaves</b>		227	1996
B Yan, Q Dai, X Liu, S Huang, Z Wang Plant and soil 179 (2), 261-268			

Listwise Collaborative Filtering

https://scholar.google.it/scholar?hl=en&authuser=1&q=Listwise+Collaborative+Filtering&btnG=&as\_sdt=1%2C5&as\_sdp=

Web Images More...

Google Listwise Collaborative Filtering

Scholar

About 1,280 results (0.09 sec)

Articles

Case law

My library

Any time

Since 2016

Since 2015

Since 2012

Custom range...

Sort by relevance

Sort by date

 include patents include citations Create alert**Listwise Collaborative Filtering**[S Huang](#), [S Wang](#), [TY Liu](#), [J Ma](#), [Z Chen](#) - Proceedings of the 38th ..., 2015 - dl.acm.org

Abstract Recently, ranking-oriented collaborative filtering (CF) algorithms have achieved great success in recommender systems. They obtained state-of-the-art performances by estimating a preference ranking of items for each user rather than estimating the absolute ...

Cited by 2 Related articles All 6 versions Cite Save More

[\[PDF\] researchgate.net](#)**List-wise learning to rank with matrix factorization for collaborative filtering**[Y Shi](#), [M Larson](#), [A Hanjalic](#) - Proceedings of the fourth ACM conference ..., 2010 - dl.acm.org

Abstract A ranking approach, ListRank-MF, is proposed for **collaborative filtering** that combines a **list-wise** learning-to-rank algorithm with matrix factorization (MF). A ranked list of items is obtained by minimizing a loss function that represents the uncertainty between ...

Cited by 99 Related articles All 16 versions Cite Save More

[\[PDF\] tudelft.nl](#)**Learning to rank: from pairwise approach to listwise approach**[Z Cao](#), [T Qin](#), [TY Liu](#), [MF Tsai](#), [H Li](#) - Proceedings of the 24th international ..., 2007 - dl.acm.org

... These include document retrieval, **collaborative filtering**, expert finding, anti web spam, sentiment analysis, and product ... metric between the corresponding top k probability distributions as the **list-wise** loss function ... when we use Cross Entropy as metric, the **listwise** loss function ...

Cited by 901 Related articles All 26 versions Cite Save More

[\[PDF\] wustl.edu](#)**Probabilistic latent preference analysis for collaborative filtering**[NN Liu](#), [M Zhao](#), [Q Yang](#) - Proceedings of the 18th ACM conference on ..., 2009 - dl.acm.org

... To recommend new items to a user, content-based **filters** match their representations to those items the user has expressed interests on. In contrast, the **collaborative filtering**(CF) approach does not require any content information about the items, it works by collecting ratings ...

Cited by 78 Related articles All 5 versions Cite Save More

[\[PDF\] cuhk.edu.hk](#)**Effort estimation based on collaborative filtering**[N Ohsugi](#), [M Tsunoda](#), [A Monden](#) - ... Conference on Product ..., 2004 - Springer

... Their results showed that **listwise** deletion technique did not performed well when the level of missing data was more than 30 ... In this paper, we propose **Collaborative Filtering** (CF) based effort estimation method, under the assumption that the (historical) predictor data have a ...

Cited by 18 Related articles All 18 versions Web of Science 2 Cite Save More

[\[PDF\] toyo.ac.jp](#)

Listwise Collaborative Filtering

https://scholar.google.it/scholar?hl=en&authuser=1&q=Listwise+Collaborative+Filtering&btnG=&as\_sdt=1%2C5&as\_sdt=

Web Images More...

Google Listwise Collaborative Filtering

Scholar About 1,280 results (0.09 sec)

Articles **Listwise Collaborative Filtering** [PDF] researchgate.net  
 S Huang, S Wang, TY Liu, J Ma, Z Chen - Proceedings of the 38th ..., 2015 - dl.acm.org  
 Abstract Recently, ranking-oriented collaborative filtering (CF) algorithms have achieved great success in recommender systems. They obtained state-of-the-art performances by estimating a preference ranking of items for each user rather than estimating the absolute ...  
 Cited by 2 Related articles All 6 versions Cite Save More

Case law

My library

Any time **List-wise learning to rank with matrix factorization for collaborative filtering** [PDF] tudelft.nl  
 Y Shi, M Larson, A Hanjalic - Proceedings of the fourth ACM conference ..., 2010 - dl.acm.org  
 Abstract A ranking approach, ListRank-MF, is proposed for collaborative filtering that combines a list-wise learning-to-rank algorithm with matrix factorization (MF). A ranked list of items is obtained by minimizing a loss function that represents the uncertainty between ...  
 Cited by 99 Related articles All 16 versions Cite Save More

Since 2016

Since 2015

Since 2012

Custom range...

Sort by relevance **Learning to rank: from pairwise approach to listwise approach** [PDF] wustl.edu  
 Z Cao, T Qin, TY Liu, MF Tsai, H Li - Proceedings of the 24th international ..., 2007 - dl.acm.org  
 ... These include document retrieval, collaborative filtering, expert finding, anti web spam, sentiment analysis, and product ... metric between the corresponding top k probability distributions as the list-wise loss function ... when we use Cross Entropy as metric, the listwise loss function ...  
 Cited by 901 Related articles All 26 versions Cite Save More

Sort by date

include patents

include citations

Create alert

**Probabilistic latent preference analysis for collaborative filtering** [PDF] cuhk.edu.hk  
 NN Liu, M Zhao, Q Yang - Proceedings of the 18th ACM conference on ..., 2009 - dl.acm.org  
 ... To recommend new items to a user, content-based filters match their representations to those items the user has expressed interests on. In contrast, the collaborative filtering(CF) approach does not require any content information about the items, it works by collecting ratings ...  
 Cited by 78 Related articles All 5 versions Cite Save More

**Effort estimation based on collaborative filtering** [PDF] toyo.ac.jp  
 N Ohsugi, M Tsunoda, A Monden - ... Conference on Product ..., 2004 - Springer  
 ... Their results showed that listwise deletion technique did not performed well when the level of missing data was more than 30 ... In this paper, we propose Collaborative Filtering (CF) based effort estimation method, under the assumption that the (historical) predictor data have a ...  
 Cited by 18 Related articles All 18 versions Web of Science 2 Cite Save More

# Data Analytics Seminar 1 -Anomaly Detection

## Finding additional material

- ▶ New e-resources are available in the university:
  - ▶ EBSCO database:
    - [https://dbis.uni-regensburg.de//detail.php?bib\\_id=ubhi&colors=&ocolors=&lett=fs&tid=0&titel\\_id=1333](https://dbis.uni-regensburg.de//detail.php?bib_id=ubhi&colors=&ocolors=&lett=fs&tid=0&titel_id=1333)
    - [https://dbis.uni-regensburg.de//detail.php?bib\\_id=ubhi&colors=&ocolors=&lett=fs&tid=0&titel\\_id=6926](https://dbis.uni-regensburg.de//detail.php?bib_id=ubhi&colors=&ocolors=&lett=fs&tid=0&titel_id=6926)
  - ▶ IEEE magazines:
    - <https://ieeexplore.ieee.org>
  - ▶ Springer e-books: [http://rzblx10.uni-regensburg.de/dbinfo/detail.php?titel\\_id=100988&bib\\_id=ub\\_s](http://rzblx10.uni-regensburg.de/dbinfo/detail.php?titel_id=100988&bib_id=ub_s)
  - ▶ Wiley E-Books:
    - <https://onlinelibrary.wiley.com/>
- ▶ Can be only used at university or via university VPN.