Titel: Trusting Data Analytics Process from the Perspective of their Stakeholders

Authors: Sven Gehrke, Sandra Niemz, Johannes Ruhland

Presenter: Sven Gehrke
Chair of Business Informatics,
Friedrich-Schiller-University,
Jena, Germany
sven.gehrke@uni-jena.de

Curiculum Vitae

Sven Gehrke
Graduate in Business Management (Diplom-Kaufmann)
* 01.04.1972

Education

1991- 1997	Business Management/ Business Informatics, Friedrich Schiller Universität Jena
1990- 1991	Mathematics / Computer Science, Friedrich Schiller Universität Jena

Professional Experience

2016 – 2022	Assistant Professor hair of Business Informatics, Friedrich-Schiller-Universität Jena, Germany
2007 – 2016	Business Consultant (ITIL, IT Governance)
1999 – 2007	Founder and CEO Information Solutions GmbH, Germany
1997 – 1999	Consultant, Peregrine Systems GmbH, Germany
1997 – 1997	Assistant Professor Chair of Business Informatics, Friedrich-Schiller-Universität Jena, Germany

Motivation

- big data analyses take up an ever larger part of our lives or influence them indirectly.
- data mining algorithms are largely based on heuristics, i.e., finding probable solutions with limited knowledge and time.
- this goes hand in hand with probabilities and trust.
- extensive literature on trust in general, but no univrsal accepted definition
- very little literature on trust in data mining projects
- common data mining methods pay little attention on this perspective

Our Approach

- comparison of major current perspectives towards trust
- investigate popular data mining methods towards trust
- identification of stakeholders in data mining projects and their interaction
- interviews with major stakeholders of the data mining process outlining open issues and challenges found during the survey

Trust from Business Informatics Perspective
Technology Acceptance Model (TAM)

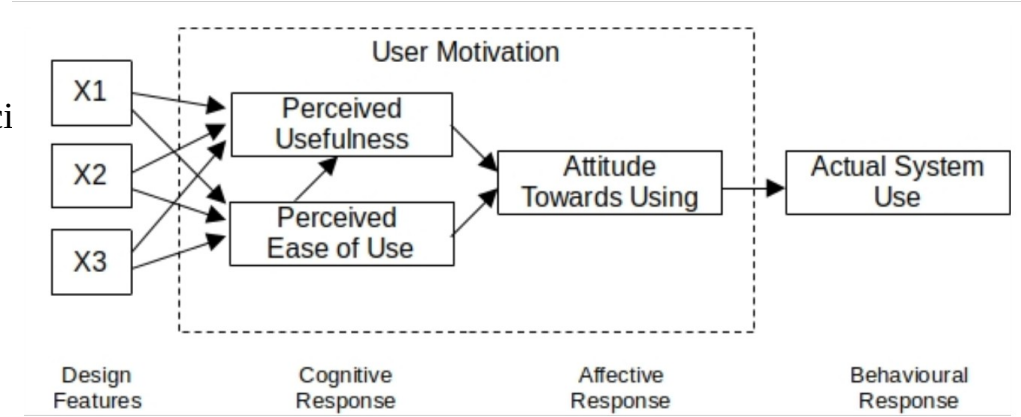
- "Attitude Towards Using" is the readiness for use
- influenced by "Perceived Usefulness" and "Perceived Ease-of-Use"
- “Perceived Usefulness” describes the expected benefit
- “Perceived Ease-of-Use” describes the costs for the user to learn how to use the technology and thus indirectly the costs of building trust.

Pro:

- due to its simplicity, it is easy to use and popular. However, the model focuses on the user and the lack of consideration of the situation is often critici

Contra:

- does not take time into account and therefore a separation of initial trust and continued trust is not described



Trust from Sociology and Psychology Perspective
Model by McKnight and Chervany

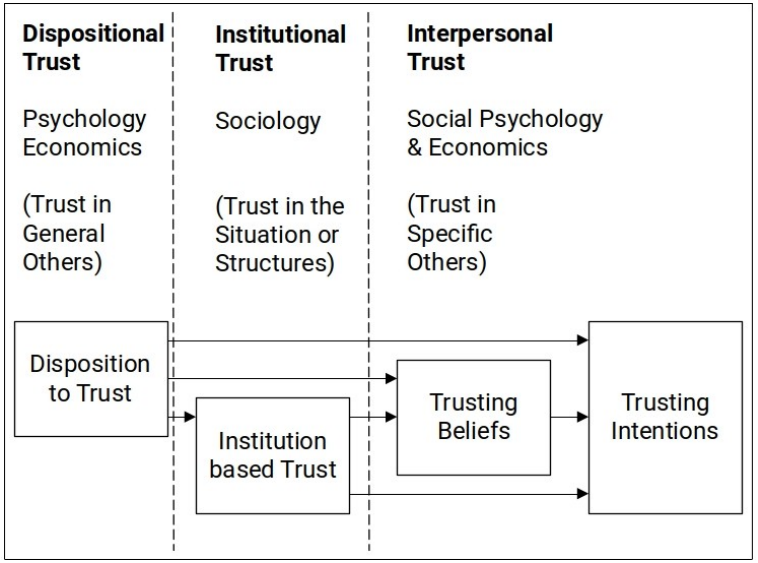
- “trusting believes” is the extent to which a target is likely to behave in a way that is "benevolent, competent, honest, predictable in a situation”
- “trusting intentions” is the extent to which a person is willing to make himself vulnerable to another persons actions
- both influenced by personal “disposition to trust” and the surrounding environments influence („institution based trust“)

Pro:

- can be applied to person and technology

Contra:

- separation of initial trust and continued trust is not described explicitly
- “continued trust” and trusting intentions result from experience and therefore the balance of incentives and penalties resulting from trusting
- “initial trust” results from trust transfer - either from person, groups or places



Trust from Microeconomic Perspective

- focus on trust in goods and the costs of evaluating their properties, less on individual disposition.
- The assumption is, that information market does not exhibit high degrees of transparency
- to evaluate the information, the information must be known therefor investments in learning and evaluation must be made
- three types of goods:
  - search goods: can be evaluated before use and therefore trusted due to previous experience or easily available information.
  - experience goods: can be evaluated only after use and therefore trusted after the use and need either a transfer of trust or reduced penalties
  - credence goods: cannot be evaluated due to prohibitive information retrieval costs or singularity and depend always on external trust transfer
- By replacing „goods“ with “data mining results“ it can be used for classifying data mining result types
- search “data mining results“ have a strong linkage towards trust transfer and/ or previous experienced trust, experience “data mining results“ need initial trust and a positive experience balance for continued usage, and search “data mining results“ cannot personally be evaluated over time at all and depend entirely on trust transfer.

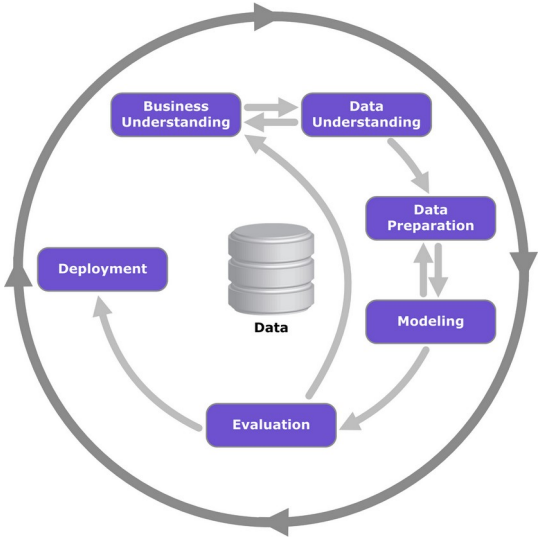
Pro:

- complementary model to TAM and the model of McKnight and Chervany

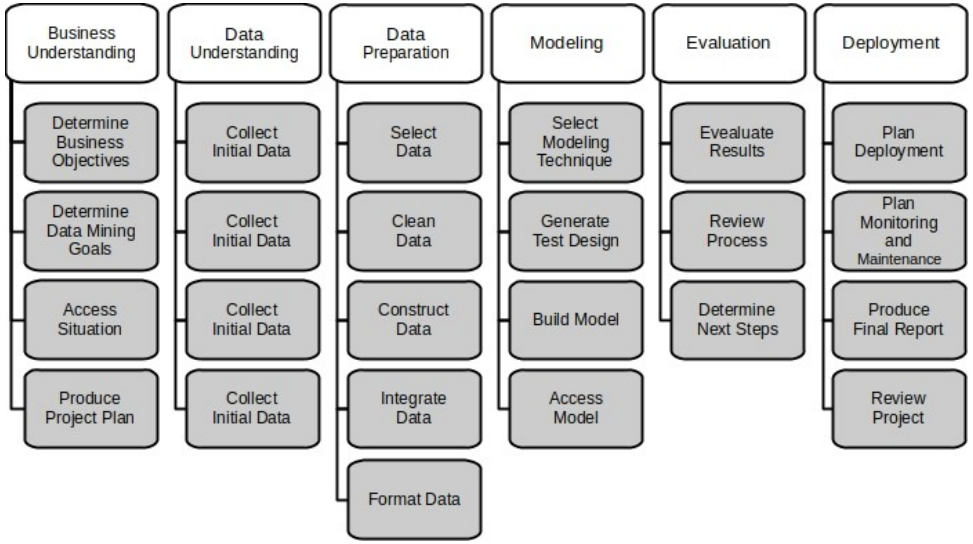
Data Mining Method CRISP DM

- no task assigned to stakeholder management or trust building

Process Stage View



Process Task View



CHANGING RESPONSIBILITIES OF STAKEHOLDERS (DESCRIBED BY RACI MATRIX)

- Responsibilties change in the stages of the CRISP DM model
- concerned persons of the deployed model are not considered yet, but needs to be considered
- In order to pass trust in the results of the previous stage to the next stage it needs „trust transfer“
- „trust transfer“ could result from trusted persons, situations or institutions (see McKnight and Chervany „Institution based trust“)

	Project Sponsor (PS)	Business User/ Analyst (BA)	Data Analyst/ Scientist (DA)	Information Ownership/ Flow
Business Understanding	**a**	**r**	c	PS → BA → DA
Data Understanding	**a**	c	**r**	BA/ DA
Data Preparation/ Modeling	**a**	i	**r**	DA
Evaluation	**a**	**r**	c	DA → BA → PS
Deployment	a	r	c	PS

a = accountable; = responsible; c = to consult; i = to inform

STAKEHOLDER INTERVIEWS

- Which factors are important from the perspective of each identified stakeholder in order to trust the data mining results?

Stakeholder	Interview Type	Interview Channel
S1: Data Analyst	semi-structured	face-to-face
S2: Business User	semi-structured	face-to-face
S3: Project Sponsor/ Management	semi-structured	face-to-face
S4A: Normal Consumer	semi-structured / closed	telephone
S4B: Informed Consumer	semi-structured	face-to-face

STAKEHOLDER INTERVIEW
S1 DATA ANALYST REPRESENTATIVE

Interview

For the interview, 3 data scientists were asked independently of each other which indicators they believe are relevant in order to trust the data and models. Then they were presented with various business KPIs and visualizations of their department together with the respective business representative and the similarities and differences in understanding were determined.

Main concerns and issues

In the business understanding, it was difficult for the stakeholders involved to interpret the specific KPIs. Concrete examples and the representation of the processes through graphics were essential for understanding.

The interviews showed that the KPIs used to justify the analysis results were rarely understood or misunderstood.

In principle, graphic representations were preferred by the other stakeholders involved. More complex representations were accepted, but required more detailed descriptions, and here again the data analysts often struggled with the business terms. As a compromise for understanding, several simple graphics that build on one another were used.

STAKEHOLDER INTERVIEW
S2 BUSINESS USER

Interview
For the interview, 3 representatives were interviewed independently of each other regarding their intentions during the phases in which they are responsible. Then they were presented with various KPIs and visualizations of their department together with the respective data analysts and the similarities and differences in understanding were determined.

Main concerns and issues
The concerns and issues of data analysts reflect the concerns and issues found among business users.

STAKEHOLDER INTERVIEW
S3 PROJECT SPONSOR/ MANAGEMENT

Interview
For the analysis, 10 senior IT managers were asked about their criteria for building trust in a guided interview. The following section summarizes the answers and the underlying intentions.

Main concerns and issues
Looking at the key considerations and underlying intentions, the focus is clearly on promoting institutional trust rather than understanding individual BDA and its metrics. The interviewees emphasized that building a high-quality and transparent data infrastructure is essential for trust in the results.

STAKEHOLDER INTERVIEW S4A CONSUMER

Interview
In a semi-structured interview, 23 people between the ages of 20 and 60 were asked which factors are relevant for them in different contexts in order to trust data mining analyses.

Main concerns and issues
The results of the data mining were accepted to a very limited extent. Without a well-founded justification for the refusal, it was doubted that the data were representative and reflected the personal circumstances. Although trust was positively influenced by the spread of the BDA (e.g., wearables/ web portals) and by certificates, the results are doubted by 70% - 80% of the respondents and used personally. When it comes to acceptance, the personal opinion of a specialist or friends prevails. In principle, the respondents do not see themselves in a position to validate the data bases and functionalities and need support from their environment.

STAKEHOLDER INTERVIEW
S4B: INFORMED CONSUMER

Interview
In the interviews, three scientists were questioned in a semi-structured interview. They were not actively involved in the analyses, but they were familiar with the environment.

Main concerns and issues
The expert survey revealed that these people generally trust the analyses, but inform themselves about the data collection, data processing and methods used on a random basis. A renowned environment of the BDA reduces the scope of own validations, but is not sufficient.

CONCLUSION

- trust in the results correlate strongly with the proximity to the process and the associated costs of information procurement.
- specialists tend to orientate themselves towards the key figures of **their** specialist area during the evaluation, which are not universally understood
- visualization seem to have a greater influence on the overall understanding than on specific key figures.

Thanks for your attention