Unstructured data
Can its power be unleashed?

Prof. dr. Jaap E. Wieringa

Report RUGCIC-2016-01
Contents

✓ Management summary ................................................................. Page 3
✓ What are unstructured data? ....................................................... Page 9
✓ Why are unstructured data relevant? ............................................ Page 15
✓ Unstructured data problems ....................................................... Page 23
✓ Examples of insights based on unstructured data ......................... Page 33
✓ Advice for working with unstructured data ................................. Page 60
✓ Resume of the author .................................................................. Page 63
✓ References .................................................................................. Page 65
✓ Customer Insights Center ............................................................ Page 73
Management summary
Background, conclusions and recommendations
What are unstructured data?

› Unstructured data are data that have no fixed data model, and are not arranged in a fixed pre-defined manner
  - Without preprocessing, unstructured data cannot be stored in a table
  - Examples: social media (tweets, blogs, posts, etc.), call center data, email, surveys with open questions, etc

› Unstructured data are strongly linked to the three V’s of ‘Big Data’:
  - Volume: unstructured data typically require more storage space than structured data
  - Velocity: the amount of unstructured data is increasing more rapidly than the amount of structured data
  - Variety: unstructured data are generated in previously untapped data sources, which may reveal very personal customer information
How prevalent are unstructured data?

- Projection: 80% of business data are unstructured data
- Unstructured data grow much faster than structured data
  - ‘Rich content’ (e.g. pictures and videos) require more data storage
  - Data produced by objects formerly not connected, such as thermostats, watches, vending machines, cars, robots, servers and heavy machinery are an important driver of data growth (‘Internet of Things’)
- Unstructured data sources become the predominant basis for customer insights
  - In combination with structured data, unstructured data sources help to obtain a more complete picture of the needs and want of customers
  - Unstructured is typically more subjective
- Structured data typically provide an answer to the ‘what’ question (e.g. what customers are likely to churn?)
- Unstructured data typically provide an answer to the ‘why’ question (e.g. why are customers churning?)
Why are unstructured data important for managers?

› Insights from unstructured data are valuable and lead to sustainable competitive advantages
  • In combination with structured data, unstructured data sources help to obtain a more complete picture of the needs and want of customers
  • Structured data typically provide an answer to the ‘what’ question (e.g. what customers are likely to churn?)
  • Unstructured data typically provide an answer to the ‘why’ question (e.g. why are customers churning?)

› But: working with unstructured data can be difficult and time consuming
  • In many organizations, tools and skills are lacking to benefit from the opportunities that unstructured data provide
  • Only a fraction of unstructured data is actually analyzed
Overview of (new) insights on unstructured data

› Insights from text analytics (examples):
  • digital word-of-mouth is central to driving business outcomes
  • a company’s press releases are surprisingly effective

› Insights from speech analytics (examples):
  • anger of inbound callers can be identified with a high level of accuracy
  • speech recognition technology can improve call center productivity

› Insights from image analytics (examples)
  • design of a product can have a stronger effect on sales performance than price
  • emotions of customers can be recognized with a high level of accuracy using facial coding

› Insights from video analytics
  • video-based automated recommendation system for use in garment retail stores can outperform benchmark recommender systems
Recommendations for marketing managers

› Unstructured data provide potentially wonderful opportunities

› However, working with these data provide time consuming and costly challenges

› Weigh the expected benefits against expected costs!

› Do not hold back: invest in the biggest opportunities

› Start each project with a relevant business question

› Invest in IT tools and IT skills of marketing analysts, but do not ignore investing in customer knowledge
What are unstructured data?

Definitions