

# From Data Mining via predictive Analysis to Surveillance capitalism

Mail: [uli@ukleemann.de](mailto:uli@ukleemann.de)

<https://ukleemann.de>

Twitter: [@hcx5](https://twitter.com/hcx5)

Blog: [b-nd.de](http://b-nd.de)

# Table of contents

- Are you aware?
- Definition data-mining predictive-analytics surveillance capitalism
- The 4 V's
- What big-data really is after professor Shoshona Zuboff
- Shoshona's 3 laws
- Business analytics vs. advanced analytics
- Advanced analytics in practice
- Descriptive modeling
- Diagnostic modeling
- Predictive modeling
- Prescriptive modeling
- Application scenarios for predictive analytics
- Users
- Chances & Risks
- Current advances
- The New Lords of the rings
- Examples for Shoshona's laws
- How surveillance capitalism works
- Why is surveillance capitalism so dangerous?
- Defense strategies

# Are you aware?

**"Most Americans are aware that there are two groups that are regularly monitored as they move around the country. The first group is monitored against their will on the basis of a court order requiring them to wear an electronic anklet. The second group includes everyone else."**

*Statement of a consultant of the motor vehicle insurance industry about the advantages of "motor vehicle telematics" and the amazingly far-reaching monitoring possibilities of these allegedly advantageous systems, which are already being used today or are in development.*

*Source: The Secrets of Surveillance Capitalism*

# Data-Mining predictive Analytics

## Surveillance Capitalism definitions:

- process of finding anomalies, patterns and correlations within large data sets to predict outcomes
- use of data, statistical algorithms and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future.
- describes a market driven process where the commodity for sale is your personal data, and the capture and production of this data relies on mass surveillance of the internet.



# What is BIG-DATA?

## 40 ZETTABYTES

[ 43 TRILLION GIGABYTES ]  
of data will be created by 2020, an increase of 300 times from 2005



## Volume SCALE OF DATA

## It's estimated that 2.5 QUINTILLION BYTES

[ 2.3 TRILLION GIGABYTES ]  
of data are created each day



Most companies in the U.S. have at least **100 TERABYTES** [ 100,000 GIGABYTES ] of data stored

**6 BILLION PEOPLE** have cell phones



WORLD POPULATION: 7 BILLION

The New York Stock Exchange captures

**1 TB OF TRADE INFORMATION** during each trading session



## Velocity ANALYSIS OF STREAMING DATA

Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure



By 2016, it is projected there will be **18.9 BILLION NETWORK CONNECTIONS**

— almost 2.5 connections per person on earth



## The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES** [ 161 BILLION GIGABYTES ]



**30 BILLION PIECES OF CONTENT** are shared on Facebook every month

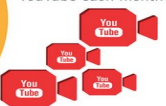


## Variety DIFFERENT FORMS OF DATA



By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

**4 BILLION+ HOURS OF VIDEO** are watched on YouTube each month



**400 MILLION TWEETS** are sent per day by about 200 million monthly active users



**1 IN 3 BUSINESS LEADERS**

don't trust the information they use to make decisions



**27% OF RESPONDENTS**

in one survey were unsure of how much of their data was inaccurate

## Veracity UNCERTAINTY OF DATA

Poor data quality costs the US economy around **\$3.1 TRILLION A YEAR**



# What Big Data really is

- Euphemism
- smuggled goods
- Stolen goods
- Fence goods

because it was acquired through a technology that we

- be unfamiliar with
- misunderstand
- we never agreed to

Source: Shoshona Zuboff

# Soshona's 3 Laws

**1<sup>st</sup> Everything that can be automated will be automated.**

**2<sup>nd</sup> everything that can be informed will be informed**

**3<sup>rd</sup> every digital application that can be used for surveillance and control will be used for surveillance and control**

**That was 30 years ago**



Prof Shoshana Zuboff

# Business Intelligence vs. Advanced Analytics

“In particular, the ability to evaluate large amounts of data (Big Data) at low cost has led to the breakthrough of Advanced Analytics in recent years. But why is AA so important?”

- data analyses do not simply examine past events, but lead to forecasts of future events
- Instead of understanding afterwards why something happened, AA puts companies in a position where they can act proactively

orientation	Looking back	Future-oriented
questions	What happened?	What will happen? If ? Next?
methods	Reporting monitoring/alerting	predictive modeling descriptive modeling
Big data	yes	yes
Data types	Structured, some unstructured	Structured and unstructured
Knowledge generation	Manual	Automatic

# Advanced Analytics in practice

- can be found in almost all industries and economic sectors
- Certain application scenarios are more obvious (risk management, churn prediction)
- can also be used to better plan the deployment of the team
- Due to the ever-increasing flood of data and its easy availability, a value-enhancing use of the data is virtually an ideal way not only to better understand the business processes, but also to better control them.

# Descriptive Modeling Step 1

**Descriptive Modeling:**

**What happened?**

**deals with the past and tries to understand the effects on the present**

# Diagnostic Modeling Step 2

Why did something happen?

provides answers to questions about the causes, effects, interactions or consequences of events

# Predictive Modeling Step 3

**What will happen? (if) / (next)?**

**looks into the future and uses data mining, machine learning and other statistical methods to predict the probability of future events**



# Prescriptive Modeling Step 4

How must we act so that a future event or does not occur?

goes one step further than predictive analytics

also provides recommendations for action on how to influence a certain trend

prevent a predicted event

or react to a future event

based on sophisticated analytical models and Monte Carlo simulations using known and random variables

displays if-then scenarios for better understanding the range of possible outcomes.

# Fazit

**In the end, you should not look at data mining as a separate, standalone entity**

**pre-processing**

**(data preparation, data exploration) and post-processing are equally essential.**

**Prescriptive modelling looks at internal and external variables and constraints to recommend one or more courses of action.**

# What are the application scenarios for predictive analytics?

**smart grid** provides load forecasts and predicts electricity demand

**scoring** banks estimate the probability or risk that a customer will not be able to make future installment payments

**maintenance** sensors transmit data on the status of a system to a cloud platform. The solution analyzes key data on usage, wear and tear and condition from various sources and thus identifies error patterns and low-quality components

# Who's using it?

- **COMMUNICATIONS** , helping them predict customer behavior and offer highly targeted and relevant campaigns.
- **INSURANCE** to price products more effectively across business lines and find new ways to offer competitive products
- **EDUCATION** With unified, data-driven views of student progress, educators can predict student performance before they set foot in the classroom
- **MANUFACTURING** can predict wear of production assets and anticipate maintenance, which can maximize uptime and keep the production line on schedule.
- **BANKING** get a better view of market risks, detect fraud faster, manage regulatory compliance obligations and get optimal returns on their marketing investments.
- **RETAIL** can offer more targeted campaigns – and find the offer that makes the biggest impact on the customer.

# Chances

## **Decision-making**

can be used to identify patterns and information hidden in the data  
eg. By evaluating machine data information can be obtained on the intervals between machine failures.

## **Increase of efficiency**

a wonderful tool for optimizing processes and reducing costs

## **eg trucking**

Predictive Analytics can perfectly optimize the routes of the trucks to save time and costs under certain circumstances

## **research and development**

existing or future trends can be identified or predicted

costs, performance or resistance to breakage could be tested by making minor changes to the product without ever having produced such a product

# Risks

## **attack on privacy**

companies that use and analyse people's data can gain deep insights into their habits and character traits. Furthermore, the protection of confidential data poses a great risk

## **Traceability**

As we move through the net we inexorably drag a trail of data behind us. Which pages we visit, which products or in which order we view them is stored (Cookies). Through the decryption and analysis of these traces, the person is connected to the data, whereby he loses any kind of anonymity

## **Criminals**

Data is the new gold.

Wherever electronic systems and virtually stored data are available, the chance of being attacked by data-thiefs is also determined

## **Incomprehensibility**

it is almost impossible to achieve perfect results using one algorithm

not all data is useful for every application and can therefore change certain results

Incorrect connections between different data can ultimately lead to wrong conclusions and decisions

# Predictive Analytics History & Current Advances

## Why now?

- **Growing volumes and types of data, and more interest in using data to produce valuable insights.**
- **Faster, cheaper computers.**
- **Easier-to-use software.**
- **Tougher economic conditions and a need for competitive differentiation.**

With interactive and easy-to-use software becoming more prevalent, predictive analytics is no longer just the domain of mathematicians and statisticians. Business analysts and line-of-business experts are using these technologies as well.

# Why is predictive analytics so important?

- **Detecting fraud**

high-performance behavioral analytics examines all actions on a network in real time to spot abnormalities that may indicate fraud, zero-day vulnerabilities and advanced persistent threats

- **Optimizing marketing campaigns**

determine customer responses or purchases, as well as promote cross-sell opportunities. Predictive models help businesses attract, retain and grow their most profitable customers.

- **Improving operations**

Airlines use predictive analytics to set ticket prices. Hotels try to predict the number of guests for any given night to maximize occupancy and increase revenue.

- **Reducing risk**

A credit score is a number generated by a predictive model that incorporates all data relevant to a person's creditworthiness



# Germans are the most skeptical, with the strongest privacy protections in the EU

- **Facebook** is growing in many countries like Indonesia or Brazil, but it has peaked in others
- **Microsoft** Authorities are "strongly dependent on a few providers at the software level "in all layers especially for Microsoft, whose products "are widely used and closely linked
- **Whatsapp** introduced an end-to-end encryption and thus became one of the most secure messengers Nevertheless, the company collects all your phone numbers and metadata. (your life)
- **Google** In one fell swoop, Google could prevent new devices in markets around the world from using the Google Play Store and other Google Play services even from regular updating  
**we're under their whip**

**the easiest way to make money was to sell  
our data on to advertisers and retailers**

*...who could track and target us to increase their sales of diapers or  
lawnmowers or diet pills*

source Shoshona Zuboff

**We owned the devices**

**they owned the servers**

**They won**

# The New Lords of the Ring

How compliant are they?

- Alphabet (Google)
- Yahoo made its case in a secret court that the broad requests were unconstitutional, and lost.
- Facebook put together teams to enable more complete cooperation with the NSA
- Microsoft
- Apple

# What we are to the Lords of the Ring?

- Neither clients nor stuff
- We are their sources
- We generate the data they are harvesting
- The Lords take and sell our data - that's their business
- The Lords sell us
- Our habits
- Our sexual orientation
- Our interests
- Our likes
- Our political or religious conversion

**All we say and all we think, where we go and who we meet – that's the price and that's for sure  
not their business, but just the beginning**

# Whats the Lords intension?

- know not only our behaviour in the past
  - know exactly how we behave here and now
  - in order to be able to influence it
- but above all**
- **want to be able to predict our future behaviour and manipulate it to their advantage**

**Everything that can be automated will be automated**



**And this is today!**

# Example 1 The smart Home

- every device is connected to the internet
- sensors record all our activities
- microphones listen to every word we say (Alexa order milk)
- Finally  
**everything is skimmed off, even our most private habits and preferences**  
**welcome to total surveillance**

everything that can be informed will be informed



**This we are today!**



# Example 2 Fitness Tracker smart watch

- collect all our vital data in real time and transfer them to a server
- allow a deeper insight into our current health situation
- allow conclusions to be drawn about our way of life

**“Thats your last beer for today, cause I know what’s good for you!”**

every digital application that can be used for surveillance and control will be used for surveillance and control



**This we all use today!**

# Example 3 the so called Smartphone

- Generates lots of so called metadata
- allows our movement radius to be precisely monitored
- enables detailed movement profiles
- can be used as a "silent bug" for unnoticed listening **effectively**

**allows comprehensive conclusions about our social environment, income, educational level, friends and acquaintances, political views and much more by evaluating the connection data**

# And this we will be tomorrow!



the guy with the whip is no longer needed everything else will be as it was

# How surveillance capitalism works



**LAVIN**

**first the lords give us something we urgently need access to information and participation**

**they act like there's no other way, like there has to be**

**we use your services in exchange for our data - we must disclose our data!**

# Why is surveillance capitalism so dangerous?



**we are already trapped in this exchange**

**it is very difficult to give up access to this information and participation**

**we shouldn't have to**

# Defense Strategies

- Any information they gather is provided by us willingly
- start paying for services which you want to use (eg. Threema, mailbox.org...)
- stop posting (send directly to your addressees)
- drop Facebook
- do not provide correct sensitive information
- Reject cookies (especially third-party cookies)
- Use open source software if possible
- User free operating systems
- Use alternates to google

**Always remember: “The only thing free is the cheese in the mouse trap.”**

# Thanks for your attention! Q&A

If you want to dive deeper

- [https://www.goodreads.com/author/quotes/710768.Shoshana\\_Zuboff](https://www.goodreads.com/author/quotes/710768.Shoshana_Zuboff)
- <https://www.faz.net/aktuell/feuilleton/the-surveillance-paradigm-be-the-friction-our-response-to-the-new-lords-of-the-ring-12241996.html?GEPC=s48>
- <https://medium.com/@tPhilosophia/shoshana-zuboff-and-nigel-thrift-on-hiding-from-surveillance-technologies-b1e7f8e850cc>
- <https://www.forbes.com/sites/cognitiveworld/2019/09/17/surveillance-capitalism-and-anti-capitalism/>
- <https://hbswk.hbs.edu/item/what-s-the-antidote-to-surveillance-capitalism>
- <https://www.faz.net/aktuell/feuilleton/debatten/the-digital-debate/shoshana-zuboff-secrets-of-surveillance-capitalism-14103616.html>

slides available at: <https://ukleemann.de/english/downloads/>