

# Legal Data Science

Bart Custers PhD MSc LLM  
Associate professor/head of research  
eLaw – Center for Law and Digital Technologies  
Leiden University

INFORM WORKSHOP  
Leiden University  
4<sup>th</sup> December 2018  
13:00-13:30



Universiteit Leiden

# Who thinks I can read your mind?



# I cannot read your thoughts, but...



weird rash|

weird rash **on big toe**  
weird rash **on arm**  
weird rash **comes and goes**  
weird rash  
weird rash **on neck**  
weird rash **on chest**  
weird rash **on face**  
weird rash **on back**  
weird rash **on hands**  
weird rash **on my arm**

Google zoeken

Ik doe een gok

# Reading your mind...

- What are you thinking when you search for...
  - Barbados
  - Prozac
  - Weird rash
  - Lolitas
  - Semiautomatic rifles...on Google?



# Why is data collected and analysed?

## Making predictions...

### ■ Examples

- Weather forecast
- Predicting stock markets
- Law enforcement
  - Predicting terrorist attacks
  - Predictive policing: surveillance
- Tax authorities
  - Finding fraud
- Corporations
  - New customers/target groups
  - Marketing new products
  - Credit scores (insurance, mortgages, loans, etc.)



# Why is data collected and analysed?

## Making predictions

### ■ Examples

- Weather forecast
- Predicting stock markets
- Law enforcement
  - Predicting terrorist attacks
  - Predictive policing: surveillance
- Tax authorities
  - Finding fraud
- Corporations
  - New customers/target groups
  - Marketing new products
  - Credit scores (insurance, mortgages, loans, etc.)



Predictions aim to assess future opportunities and risks.

# Applying data science to law

Topic of this presentation:

- Data science as a new research method...

... in the legal domain

How can legal research benefit from data sciences?



This presentation is NOT about regulating data sciences or big data.

# Disrupting law school

## DISRUPTING LAW SCHOOL:

How disruptive innovation  
will revolutionize the  
legal world

BY MICHELE R. PISTONE AND MICHAEL B. HORN

CLAYTON CHRISTENSEN INSTITUTE  
for DISRUPTIVE INNOVATION



- Disrupting law school...
  - In the USA, it is less desirable to be a lawyer today.
  - Law school applications are down 40 % since 2005.
  - Employment rates for law graduates are shrinking
  
- Advantages in technology and business model innovations are changing the traditional legal model:
  - Legal solutions are more standardized and systematized
  - Lawyers become more productive due to technological innovations
  - Technological innovations offer solutions for non-lawyers

# Data science in legal research

## New perspectives

- Legal documents (legislation, case law, policy documents, research papers, etc.) are data:
  - increasingly available in electronic formats
  - new data analyses technologies (viz. big data analytics), such as data mining and text mining
  - increased accessibility of files that were previously only available on paper
- New research methods in the legal domain that are more data-driven.
  - Focus on statistical relations rather than causal relations
  - Less focus on hypotheses and theory
  - Increased efficiency and effectiveness
  - Novel, unexpected results



# Data science in legal research

Data science may contribute to answer different types of legal research questions

	What is existing positive law	How to interpret existing positive law	Sociology of law	Developing legal theory	Developing improved legislation
Quantitative legal predictions	X	X	X	X	
Accelerate research	X	X	X	X	X
Improved regulation			X	X	X



# Quantitative legal predictions

- Ruger et al. (2004): tournament to predict US supreme court rulings
  - Individual justices:
    - Human experts: 68 %
    - Model: 67 %
  - Case outcome
    - Human experts: 59 %
    - Model: 75 %
- Katz et al. (2014)
  - Model independent of
    - Justice transition
    - Ideological shift of justices
    - Macro political/economic conditions



## Predicting the Behavior of the Supreme Court of the United States: A General Approach

Daniel Martin Katz<sup>1</sup>, Michael J Bommarito II<sup>2</sup>, and Josh Blackman<sup>3</sup>

<sup>1</sup>Michigan State University - College of Law

<sup>2</sup>Michigan State University - College of Law

<sup>3</sup>South Texas College of Law

### ABSTRACT

Building upon developments in theoretical and applied machine learning, as well as the efforts of various scholars including Guimera and Sales-Pardo (2011), Ruger et al (2004), and Martin et al (2004), we construct a model designed to predict the voting behavior of the Supreme Court of the United States. Using the extremely randomized tree method first proposed in Geurts et al (2006), a method similar to the random forest approach developed in Breiman (2001), as well as novel feature engineering, we predict more than sixty years of decisions by the Supreme Court of the United States (1953-2013). Using only data available prior to the date of decision, our model correctly identifies 69.7% of the Court's overall affirm / reverse decisions and correctly forecasts 70.9% of the votes of individual justices across 7,700 cases and more than 68,000 justice votes. Our performance is consistent with the general level of prediction offered by prior scholars. However, our model is distinctive as it is the first robust, generalized, and fully predictive model of Supreme Court voting behavior offered to date. Our model predicts six decades of behavior of thirty Justices appointed by thirteen Presidents. With a more sound methodological foundation, our results represent a major advance for the science of quantitative legal prediction and portend a range of other potential applications, such as those described in Katz (2013).

Keywords: Supreme Court, Machine Learning, Law & Social Science, Quantitative Legal Prediction

### Data for the Katz-model:

- Over 60 years
- 7,700 cases
- 68,000 justice votes

# Accelerate legal research



## ■ Ravellaw

<https://www.ravellaw.com/search/m?query=privacy>

## ■ ROSS

<https://www.youtube.com/watch?v=sbwEO8-Voxo>  
(1'05'')



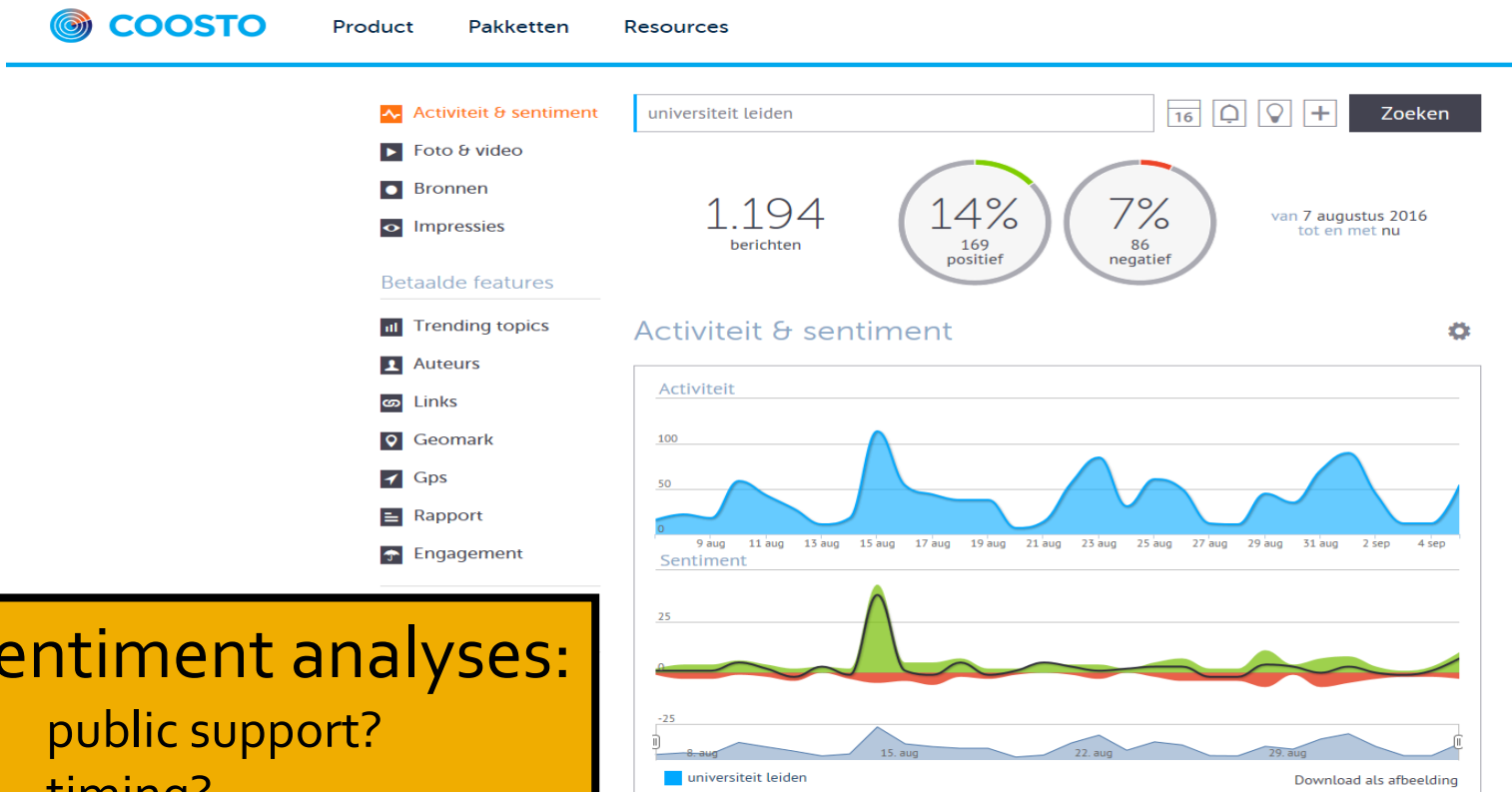
### Advantages

- Decreasing the likelihood of missing important information
- Increasing accuracy and reliability of research
- Potential of novel, unexpected results

# Improved regulation – public acceptance

Not only legal documents are relevant data:  
example of social media

COOSTO: <http://open.coosto.com/nl/>



## Sentiment analyses:

- public support?
- timing?

# Improved regulation - enforcement

## ■ Nudging:

Use social media data for improved enforcement

- Increase knowledge about how people will behaviorally respond to proposed legislation
- This may help to create legislation that can be enforced more easily...
- ... by nudging people into the expected behavior



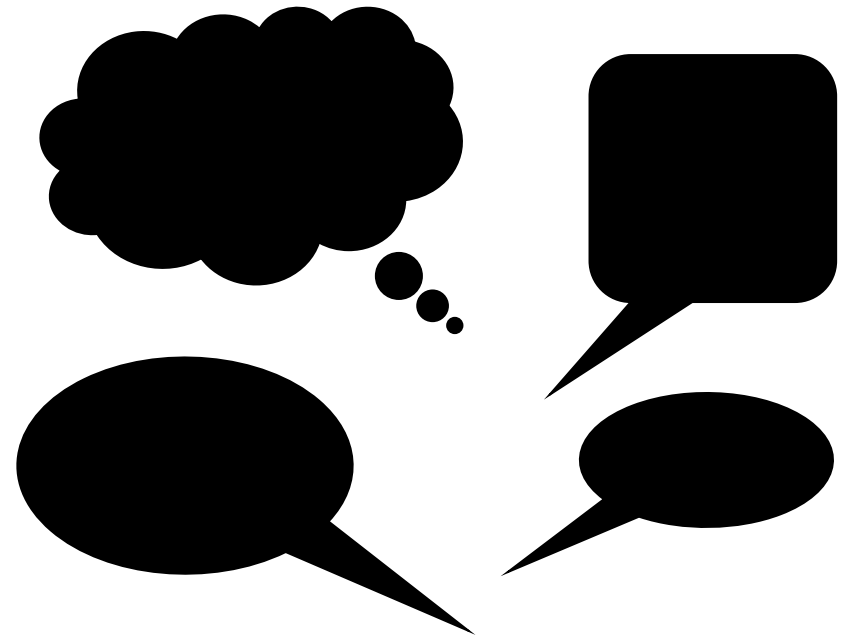
# Conclusion

Data science does not replace or invalidate other research methods

- Data science may be complementary to other methods:
  - It may disclose **novel, unexpected results**
  - It may **improve efficiency/effectiveness** of research
- Data science, however, provides limited opportunities for:
  - Investigating causality
  - Deeper understanding
  - Theory development



# Questions?



Thank you for your attention!

Or contact me later: [b.h.m.custers@law.leidenuniv.nl](mailto:b.h.m.custers@law.leidenuniv.nl)