Legal Implications of Data Mining

Liane Colonna
The Swedish Law and Informatics Research Institute
Overview

1. Introduction to the research problem
2. Data mining and EU data protection principles
3. US surveillance
4. Applicable law and jurisdiction
5. Some general conclusions about EU/US law
Part 1: Introduction to the research problem
Data mining is a dynamic form of data processing.

It can make sense of huge amounts of data.

It creates “new” knowledge.

It can generate hypotheses.
High-level goal

Verification (deductive, hypothesis driven, top down, analytic)

Tasks
- Query and reporting
- Multidimensional analysis such as Online Analytical Processing (OLAP)
- Statistical analysis

Discovery (inductive, data-driven, bottom up, synthetic)

Tasks
- Predictive/supervised
- Descriptive/Unsupervised

Predictive/supervised

Tasks
- Classification
- Regression
- Other

Descriptive/Unsupervised

Tasks
- Clustering
- Association
- Other

Classification

Technique/algorithm
- Decision tree
- Linear regression
- Other

Regression

Technique/algorithm
- Linear regression
- Other

Other

Clustering

Technique/algorithm
- K-means
- Other

Association

Technique/algorithm
- Bayesian networks
- Other

Other
Ways terrorist use the Internet

Communicate
Recruit
Fundraise
Train
Launch propaganda videos
Disseminate racist and xenophobic material
Threaten attacks
Plan attacks
Implement attacks

Ways data mining can be used to identify terrorist connections
A chilling effect on innocent behavior/threats to autonomy?

Potential for abuse or misuse such as by government bureaucrats?

A slippery slope towards a society smothered under a veil of constant surveillance?

A loss of control over data with real-world consequences?
Overall Objectives

To assess whether data mining can be reconciled with core EU data protection principles such as purpose limitation, data minimization and data accuracy.

To explore whether the legal regulation of data mining in the US national intelligence context is sufficient from a European perspective.

To examine the jurisdictional issues that invariably arise in the context of regulating national intelligence surveillance which, in the digital age, is inherently transnational.
Part II: Data mining and the core EU data protection principles
-- the purpose limitation principle--

• Article 5(b) CoE Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data
  – "Personal data undergoing automatic processing shall be ... stored for specified and legitimate purposes and not used in a way incompatible with those purposes"

• Article 6(1)(b) of EU Data Protection Directive 95/46/EC
  – “Personal data must be “collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes.”

• Article 3 of EU Framework Decision 2008/977/JHA
  – “Personal data may be collected by the competent authorities only for specified, explicit and legitimate purposes in the framework of their tasks and may be processed only for the same purpose for which data were collected…”
“What I need is an exact list of specific unknown problems we might encounter.”
Is Data Mining Voodoo Science?

The Scientific Method: Hypothesis → Data

• Develop a hypothesis based on sound theory
• Collect the data
• Test your hypothesis

The Data-Mining Method: Data → Hypothesis

• Get the data from somewhere, anywhere (and lots of it!!)
• Select a data-mining technique that seems applicable for your dataset
• Apply the technique to the data and see what turns up
• Evaluate the reliability of the automatically generated hypothesis
For what specific and legitimate purpose will you be processing my data for?

I can’t tell you yet. First, I need to mine your data and then I will know for what purpose to use the data.
DM Causes an Obfuscation of the PL Principle

Instead of stating a clear and specific purpose, the promises of DM incentivizes data controllers to create broadly defined purposes in order to bolster their freedom to collect and process data.
DM Creates “New” Information that is Hard to Control !!!
DM creates fragments of data that may not be obviously related to any particular data subject.
So, how can we reconcile DM and PL?

- Misuse model?
- Privacy by design?
- Accountability?
You give data mining a bad name!
Case Study Conclusion
The main substantive limitation imposed is that the acquisition of data must be for the purpose of obtaining foreign intelligence information.
No 4th Amendment protections

No statutory safeguards (e.g. targeting and minimization of data collection)

No independent agency that is empowered to investigate claims of wrongful surveillance

Only have standing to bring claims in US courts in a very narrow situation (e.g. private claim for willful violation of FISA)

Non-US persons have a lack of rights concerning the information collected about them
Data Mining Program

- Must precisely describe the scope of the State’s authority
- Must make sure citizens can foresee DM program
- Must set forth procedures to limit State’s authority
- Define the offences for which surveillance measures are permitted
- Define the categories of persons who may become subject to such measures
- Place limits the duration of surveillance
- Have procedures for examining, using, and storing data from the surveillance
- Ensure precautions are taken when communicating data to other parties
- Specify the circumstances for the destruction of recorded information from the surveillance
- Notification about the processing provided to the data subject

- Ensure individuals have a right to seek redress from an independent agency.
- Must make sure the program is effective in achieving its aim? Review empirical evidence?
- Must publish the regulation and avoid internal set of rules.
- Must make sure the DM program is explicitly authorized by domestic law
- Must make sure the program is explicitly authorized by domestic law.
Part IV: Applicable law and jurisdiction
Article 4(1)(a) of European Data Protection Directive

1. Which exact processing of personal data has to be examined?

2. Is there an “establishment” of the controller in the EU? (look to ECJ case law)

3. If there is an establishment, does the personal data processing in question take place “in the context of the activities” of that establishment?

(A) Look at the kind of data processing that is going on (regardless of where in the world it is happening) and (B) look at the establishment and ask whether it has a relevant role in the particular data processing in question.

If no, check to see whether Article 4(1)(c) applies.

If yes, the EU law applies.

Requires that “both human and technical resources necessary for the provision of particular services are permanently available” branch=yes, server=no
Google Spain (Case C-131/12)
Is Google, in its capacity as an online search engine, required to block unwanted personal data appearing in search results?

Q1. Which exact processing of personal data has to be examined?
- Data processing relating to search engines must be examined

Q2. Is there an establishment of the controller in the EU? – YES! Google has an establishment in Spain, Google Spain SL set up solely for marketing and advertising purposes

Q3. If there is an establishment, does the personal data processing in question take place “in the context of the activities” of that establishment?

Google’s opinion: No. Google Spain only processes data for marketing purposes – i.e. it has no control over the particular data processing at issue in the case – data processing associated with the search engine operations which, according to Google, all happens in California.

AG’s opinion: The personal data processing in question (search engine processing) is carried out “in the context of the activities” of Google Spain because Google Spain bridges too (interrelated limbs of the business).

CJEU’s opinion: The advertising activities of Google Spain were "inextricably linked" to Google Inc.'s search engine: without the sales and promotion of the online advertising space, the search engine would not be economically profitable.
Extraterritoriality and applicable law provisions in the General Data Protection Regulation

- Establishment
- Offering goods or services
- Monitoring behavior
Conclusions about applicable law and jurisdiction

• Enforceability of provision is unlikely?
• The adequacy requirement is obscured and an overly complex regulatory environment is created which may hamper international data flows?
• A de facto global standard is created?
• The concept of data interoperability is hampered?
Thank You!!
liane.colonna@juridicum.su.se