Data anonymization: becoming GDPR compliant in the era of AI and Big Data

Inés Ortega Fernández
Technical Manager of Data Analytics & AI | Security & Privacy Department
Outline

1. GRADIANT
2. INFINITECH H2020
3. Data anonymization & GDPR
4. Data anonymization in Big Data
5. GRADIANT @ INFINITECH H2020
GRADIANT
Since 2008, focused on technological development and knowledge transfer to industry

135
Professionals

6M€
Revenue in 2021

68%
Contracted companies

32%
Competitive public funding

20+
European projects

2
Cervera network projects
INFINITECH H2020

Tailored IoT & BigData Sandboxes and Testbeds for Smart, Autonomous and Personalized Services in the European Finance and Insurance Services Ecosystem
Data Anonymization for GDPR compliance

General Data Protection Regulation

- In place since May 2018
- Security & Privacy by Design
- Anonymization allows to process data without having to follow the GDPR

Data Anonymization

- Transforming personal data to irreversibly prevent identification
- Robustness of anonymization techniques
  - Possible to identify a single person
  - Link different records regarding the same individual
  - Quantity of information that can be inferred regarding the data subject

Anonymization Operations

Randomization

- Noise addition
- Permutation techniques
- Differential privacy

Generalisation

- k-anonymity
- l-diversity
- t-closeness
The privacy / utility trade-off

Data Anonymization in Big Data: challenges

- Increment in data volume
- Data variety
- Classical anonymisation techniques such as k-anonymity need to be adapted
- Real-time constraints

GRADIANT @ INFINITECH H2020

- We are designing and developing an advanced *data anonymization* tool

- Analyze + Anonymize: *risk-driven approach*
  Help customers find the best set of anonymization operations to fit their privacy and utility goals

- Data anonymization in disruptive scenarios
  - Connected vehicles (GPS data anonymization)
  - Health insurance
Inés Ortega Fernández
iortega@gradiant.org