

DATI PERSONALI E INFORMAZIONI SENSIBILI: TECNICHE DI DATA ANONYMIZATION IN R



Giovanna de Vincenzo

Founder RLadies Bari, Data Scientist InfoCamere

AGENDA



- Mi presento...
- Anonimizzazione e pseudo
- Identificatori
- Anonimizzazione: perchè è difficile?
- Privacy differenziale
- Caso d'uso
- Vi presento...RLadies Italy

A presto!

Giovanna de Vincenzo

Data Scientist @
InfoCamere

Founder
RLadies Bari

Co-founders **RLadies
Italy**

RLadies

Data Viz Lover



Repo

GitHub
[giodev11/DPI-Novem
bre2020](https://github.com/giodev11/DPI-Novembre2020)

Here I am



Contact
giodevince@gmail.com



[giodevincenzo](https://twitter.com/giodevincenzo)

PERCHÉ LA DATA PRIVACY E L'ANONIMIZZAZIONE

Il mio punto di vista

ANONIMIZZAZIONE E PSEUDONIMIZZAZIONE

Pseudonimizzazione

[...] trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'ausilio di informazioni aggiuntive [...] GDPR – Articolo 4 comma 5

Un dato pseudonimizzato può correre il rischio di essere ricostruito

Anonimizzazione

Sostituzione permanente del dato personale con una stringa senza alcun valore e a carattere assolutamente generico



ESEMPIO

1. Medico del paziente
2. Utente reparto statistiche
3. Utenti non autorizzati
4. Open data

Nome	Giulia Bianchi
Genere	F
Diagnosi	Febbre emorragica

Schermata con dati in chiaro

Nome	Silvia Rossi
Genere	F
Diagnosi	Febbre emorragica

Schermata schermata con dati pseudonimizzati

Nome	No Access
Genere	F
Diagnosi	Febbre emorragica

Schermata dati non accessibili

Nome	XXX
Genere	F
Diagnosi	Febbre emorragica

Schermata con dati anonimi

PSEUDONIMIZZAZIONE: IMPLICAZIONI



1

Rischio **interno** all'azienda

2

Rischio **esterno** all'azienda

3

Natura del dato

OLD BUT GOLD

NETFLIX



BIG DATA, SICUREZZA E PRIVACY: LA SFIDA EUROPEA DEL GDPR

COME POSSIAMO AVERE LA
CERTENZA CHE UN DATASET
SIA REALMENTE
ANONIMIZZATO?

IDENTIFICATORI DIRETTI E INDIRETTI

Identificatori diretti

1. nome,
2. cognome,
3. codice fiscale
4. ...

Identificatori indiretti

1. tragitto giornaliero
casa - ufficio
2. modello dello smartphone
3. modello dell'auto
4. applicazione per il
traffico
5. le connessioni sociali,
6. lo storico medico
7. la lingua parlata

IDENTIFICATORI DIRETTI E INDIRETTI

Identificatori diretti

1. nome,
2. cognome,
3. codice fiscale
4. ...

Identificatori indiretti

1. tragitto giornaliero
casa - ufficio
2. modello dello smartphone
3. modello dell'auto
4. applicazione per il
traffico

**La combinazione di queste informazioni,
con le quali è possibile distinguere gruppi
limitati di persone, ti rende unico**

IDENTIFICATORI DIRETTI E INDIRETTI

Identificatori diretti

1. nome,
2. cognome,
3. codice fiscale
4. ...

Il problema è reale poiché gli identificatori non sono facili da individuare in un **dataset**

Rischio: de-anonimizzazione dataset

Identificatori indiretti

1. tragitto giornaliero
casa - ufficio
2. modello dello smartphone
3. modello dell'auto
4. applicazione per il
traffico

La combinazione di queste informazioni, con le quali è possibile distinguere gruppi limitati di persone, ti rende unico

ANONIMIZZAZIONE DATI PERSONALI: PERCHÉ È DIFFICILE?

Tecnica k-anonymity: assicura che ciascun attributo descriva una popolazione di almeno k individui

[Difficoltà] per assicurare una **corretta anonimizzazione:**

1. Robustezza contro la **re-identificazione**
2. **Ipergeneralizzazione** analisi
3. Non esiste un processo standard per l'**anonimizzazione dati personali**

ANONIMIZZAZIONE DATI PERSONALI: PERCHÉ È DIFFICILE?

Margine
Operativo

1. Conoscere le debolezze di ciascuna tecnica di de-identificazione (anonimo-noto)
2. Machine learning permette di spostare il problema privacy dai dati ai modelli. E' possibile applicare tecniche di **differential privacy** all'output del modello anziché sui dati

INTRODUZIONE: DIFFERENTIAL PRIVACY

“Qual è la massima quantità di informazioni private che renderò pubbliche se rivelassi questa osservazione?”

USE CASE

TECNICHE DI

ANONIMIZZAZIONE IN R

A photograph showing three hands of different skin tones (light, medium, and dark) stacked on a wooden table. The hands are positioned in a way that suggests support or teamwork. In the background, a white smartphone lies on the table, displaying the time 10:24. A notebook with a diagram is also visible. The text 'R-Ladies Italy' is overlaid in the bottom right corner.

R-Ladies Italy

R-LADIES Mission



Gabriela de Queiroz
San Francisco, 2012

1. Promuovere la **diversità di genere** in un campo che ancora oggi è fortemente maschile
2. Focus su **R**, linguaggio di programmazione / analisi statistica
3. Opportunità di **mentorship, apprendimento e supporto**
4. Sentimento di appartenenza ad una **community**

La mission è semplice

Perché siamo qui

- Leadership femminile
- Far parte della grande famiglia di useRs e RLadiers
- Unire appassionati di data science & ML
- Favorire contesto innovativo, nuove idee e partnership

I nostri obiettivi

- Meetup e Webinar
- Facilitare l'apprendimento di R
- Sostenere e tutorare le donne nella crescita professionale nel settore
- Essere un punto di riferimento locale



R-LADIES ITALY

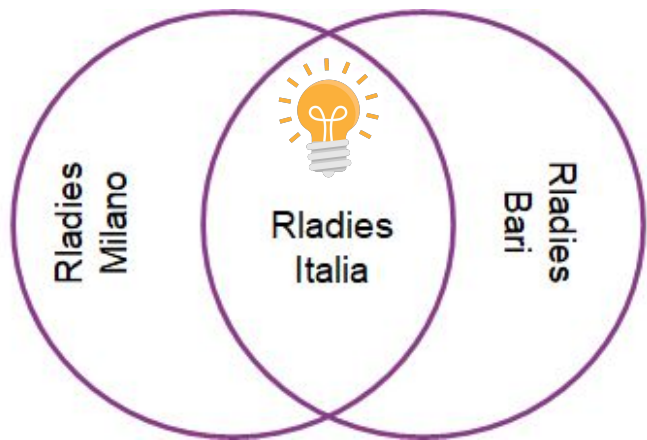
WORK IN
PROGRESS

PROGETTO #1

Il progetto si realizza attraverso una **serie di webinar** pensati come strumento che permetta di **navigare il mare di opportunità professionali** legate al settore.

Con l'obiettivo di offrire una visione ampia e completa, dedichiamo particolare attenzione **soprattutto a ruoli in settori non tradizionalmente legati alla Data Science**, e.g. HR.

Le **donne** sono le protagoniste dei nostri webinar.



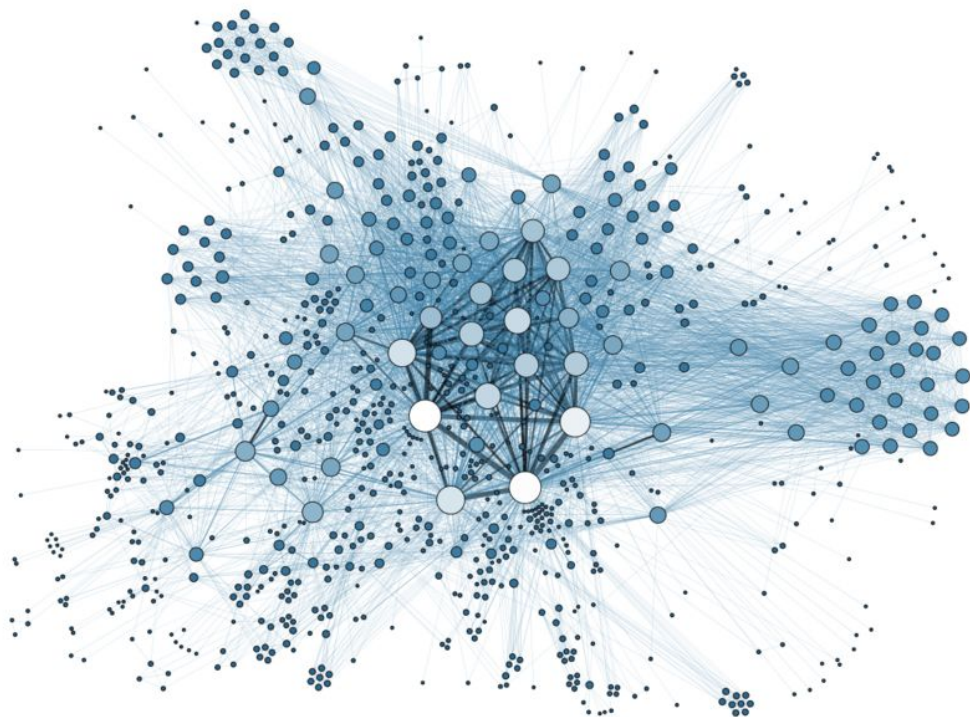
Progetto nato a luglio 2020 a fronte del lock-down

COSA PROPONIAMO

- Supporto nella fase di start-up di **altri chapters** di Rladies in **Italia**
- **Collaborare** con altri Rladies a livello mondiale
- Nuove **partnership** con altri gruppi di Data Science
- Collaborare con altre aziende sul **Diversity&Inclusion**
- **Hackathon**
- **Technical** workshops
- R [SatRday](#) Conference Italy
- **Mentorship**
- Newsletter & site dedicato



| BUILDING A NETWORK



- Partner con

/DATA



acea
energia



NUOVO WEBSITE!

Scannerizza il QR code per
accedere al nostro sito ed
essere sempre aggiornat@
sulle notizie!



Restiamo in contatto!



[meetup.com/R-Ladies-Bari/](https://www.meetup.com/R-Ladies-Bari/)



facebook.com/Rladiesbari/



twitter.com/RladiesBari



linkedin.com/r-ladies-bari/



[meetup.com/R-Ladies-Milan/](https://www.meetup.com/R-Ladies-Milan/)

facebook.com/Rladiesmilan/

twitter.com/RladiesMilan

linkedin.com/Rladiesmilan/



Rladiesitaly.slack.com

- Job opportunity
- Discussion
- Events

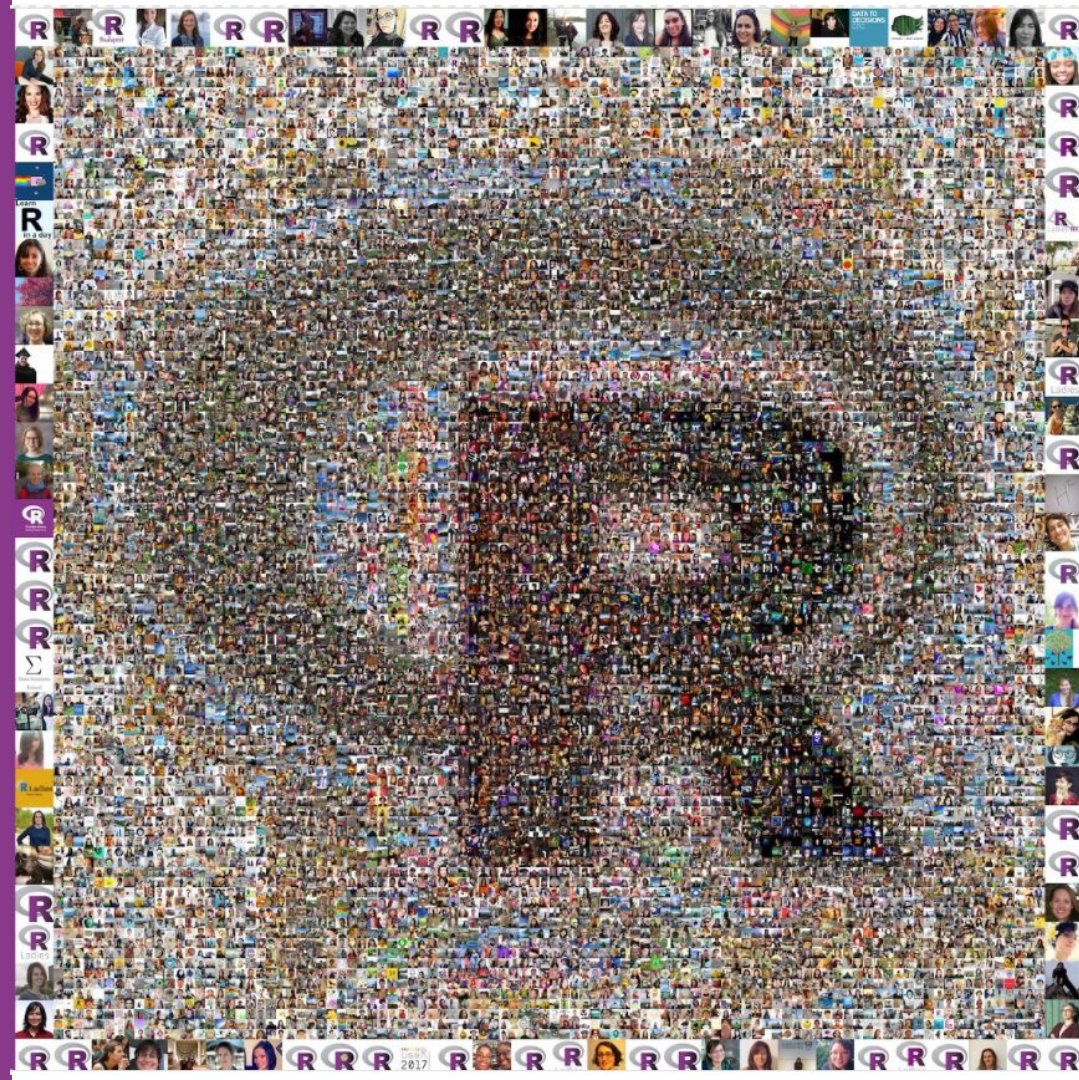


linkedin.com/giodevincenzo

www.saraiacozza.com



linkedin.com/parvaneh-shafiei-867bb73b



Be Ready
foR the
jouRney!