



Un viaggio nei dati e negli algoritmi

Cerved

Imprese e matematica

@Dipartimento di Matematica Università di Bologna



Cerved, la data-driven company



> **3.000**
PROFESSIONISTI



> **1,1 PETABYTE**
DI DATI ARCHIVIATI NEI NOSTRI 1.000 SERVER



> **1.000**
RICHIESTE DI INFORMAZIONI AL
MINUTO SU 5,6 MILIONI DI AZIENDE
ATTIVE



> **40M**
DI INVESTIMENTI STANZIATI OGNI ANNO
PER AGGIORNARE E ARRICCHIRE I
DATABASE CON NUOVI DATI E FONTI
INFORMATIVE



*Aiutiamo il Sistema Paese
a proteggersi dal rischio
e a crescere in modo sostenibile.*

I nostri elementi distintivi

1

ECOSISTEMA DATI

- ⌋ Il più grande ecosistema di dati in Italia, integrabile con quello dei nostri partner e dei nostri clienti
- ⌋ Patrimonio informativo ricco, unico, distintivo e costantemente in crescita



2

TECNOLOGIE DI FRONTIERA

- ⌋ Le Tecnologie più innovative ed avanzate per la gestione, elaborazione e visualizzazione dei dati
- ⌋ Machine Learning, Advanced Analytics e Motori Semantici per trasformare i dati in soluzioni di valore



3

APP VALUE BASED

- ⌋ «App» pronte all'uso, che non necessitano di attività di integrazione sui sistemi (IT Independent)
- ⌋ App Cerved per generare valore



I nostri elementi distintivi

1

ECOSISTEMA DATI

- ⌋ Il più grande ecosistema di dati in Italia, integrabile con quello dei nostri partner e dei nostri clienti
- ⌋ Patrimonio informativo ricco, unico, distintivo e costantemente in crescita

2

TECNOLOGIE DI FRONTIERA

- ⌋ Le Tecnologie più innovative ed avanzate per la gestione, elaborazione e visualizzazione dei dati
- ⌋ Machine Learning, Advanced Analytics e Motori Semantici per trasformare i dati in soluzioni di valore

3

APP VALUE BASED

- ⌋ «App» pronte all'uso, che non necessitano di attività di integrazione sui sistemi (IT Independent)
- ⌋ App Cerved per generare valore



Istituzioni locali e internazionali guardano a Cerved come punto di riferimento per l'economia italiana

I nostri clienti

 **95%**
delle banche italiane

 **+30.000**
aziende¹

 **+150**
Pubbliche Autorità

1. Comprese le micro aziende con acquisti su cervedirect.com

Istituzioni
primarie
che utilizzano
i nostri dati



osservatore chiave sulle PMI
per le consultazioni periodiche
previste dall'articolo IV
(Country Surveillance)



Fornitore di dati sulle
procedure concorsuali
e sui termini di
pagamento delle
aziende



BANCA D'ITALIA
EUROSISTEMA

Fornitore di indicatore sintetico
del rischio affrontato dalle
aziende e istituzioni finanziarie

ABI Associazione
Bancaria
Italiana

Fornitore di dati sulle
procedure concorsuali e sui
termini di pagamento delle
aziende

Ecosistema di dati unico



Chi siamo

3 Data Scientist, 3 anime diverse di Cerved



Michele Voto
Risk Analytics
Innovation



Roberto Santoro
Text and Graph
Analytics



Gabriele Sotto
GeoData and NLP



Data science dal punto di vista di un matematico

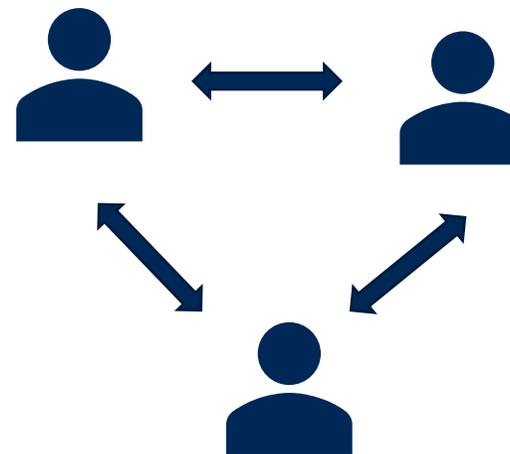
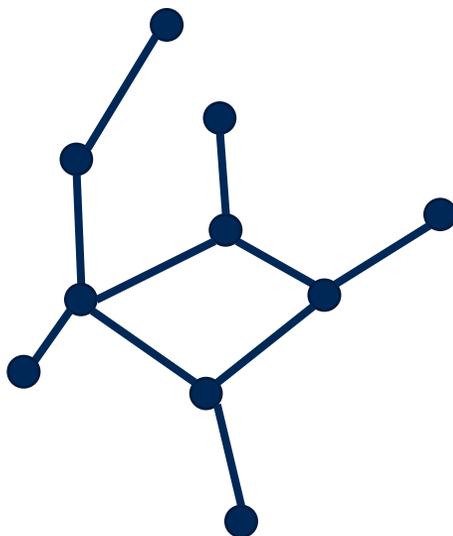
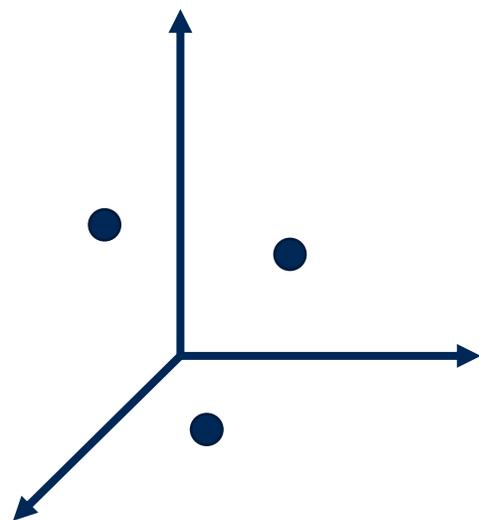
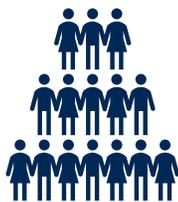
Data science: dai dati al valore



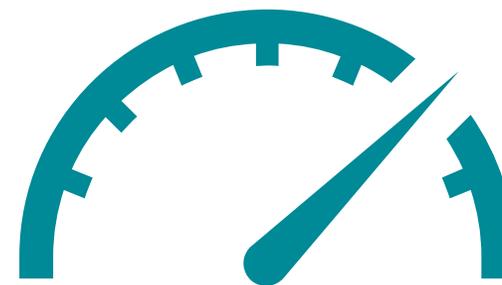
- Esperti di dominio
- Analisi statistica
- Machine Learning e AI



Data science: dai dati al valore



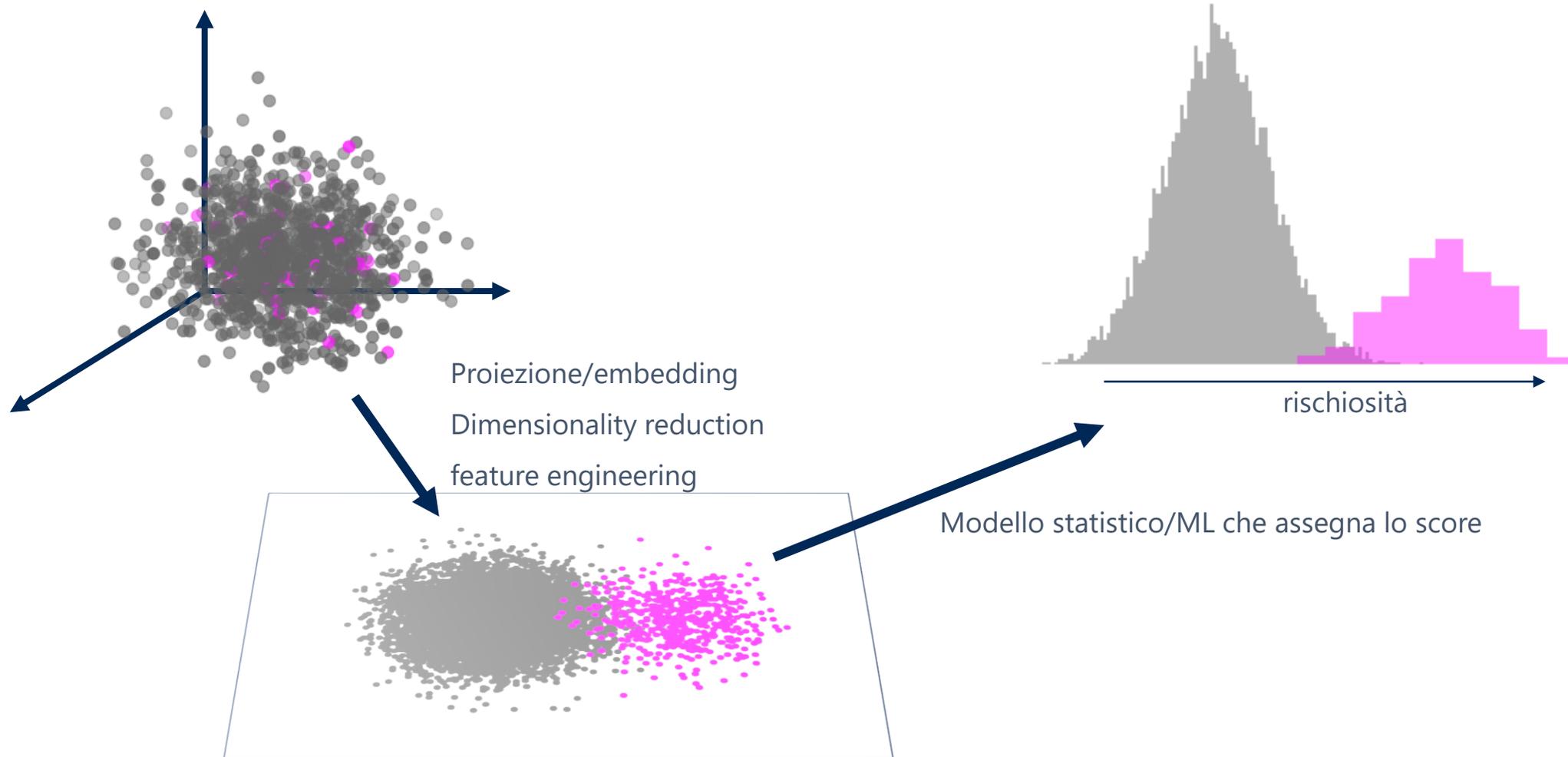
Modelli di rischio sulle aziende



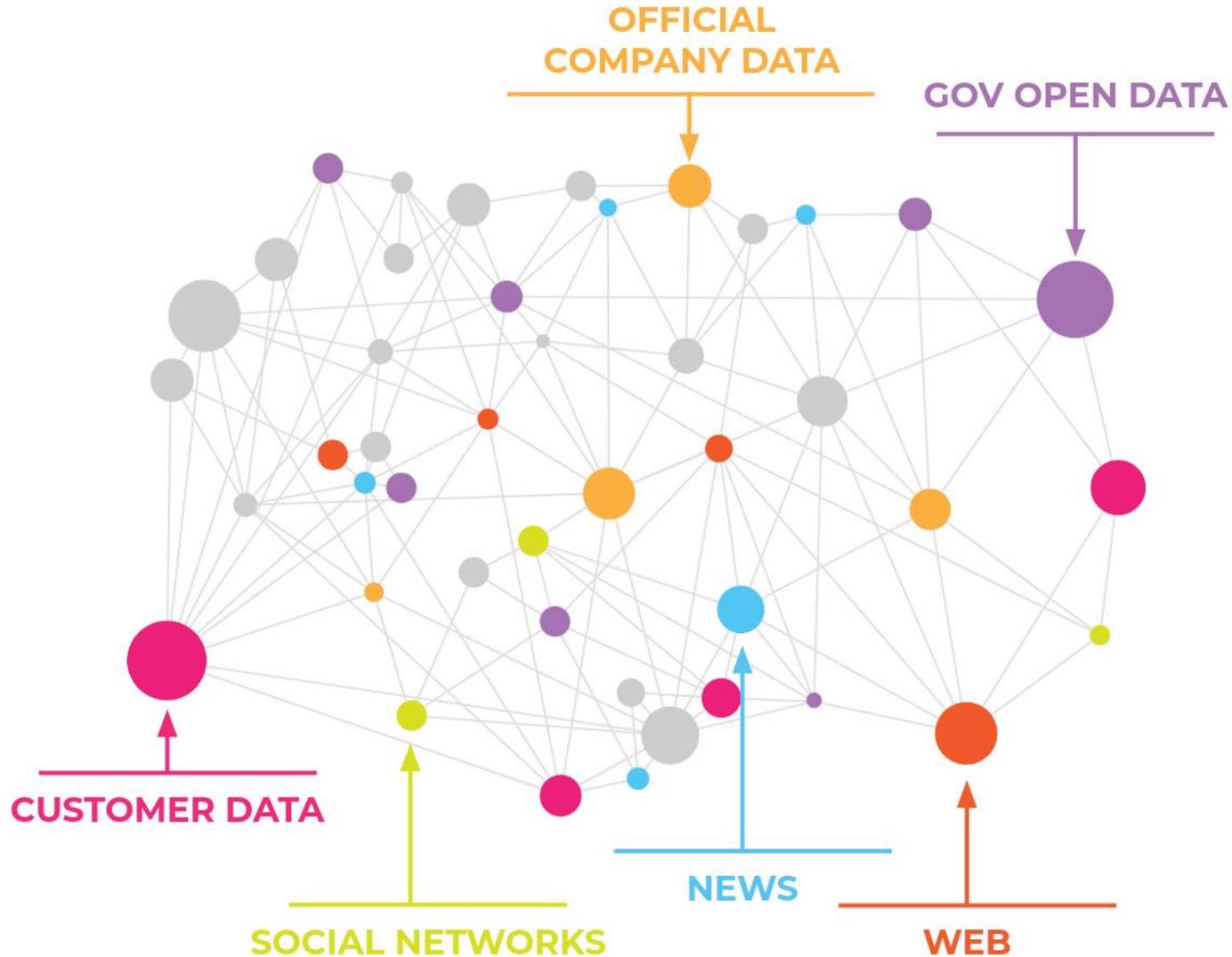
Probabilità che l'azienda sia insolvente

- Settore: costruzioni
- Ricavi 2021: 4M€
- Debiti finanziari 2021: 300k€
- 2 capannoni, 1 magazzino
- Fatture pagate ...
- Fatture da riscuotere ...
- Dati di C/C

Modelli di rischio: l'astrazione



Fonti Big Data



- 6M aziende
- 15M ruoli (persone)
- 1M siti web aziendali
- 13K news al giorno

Data Pipeline

- Ingestion
- Curation
- Linking
- Enrichment (Text Analytics, etc.)

Grafi

$$G = (V, E), E \subseteq \{(u, v): u, v \in V\}$$

Modellano:

Strade

Reti di computer

Social networks

...

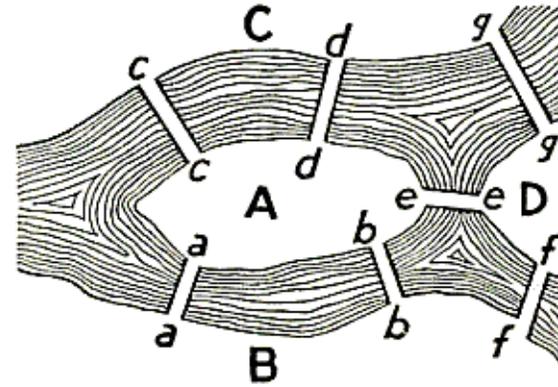
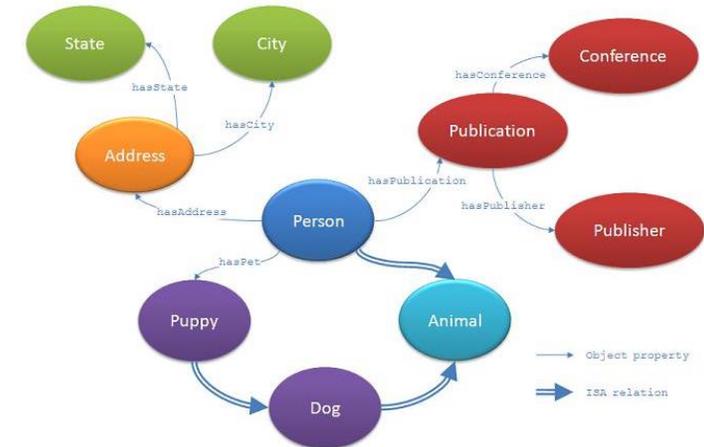
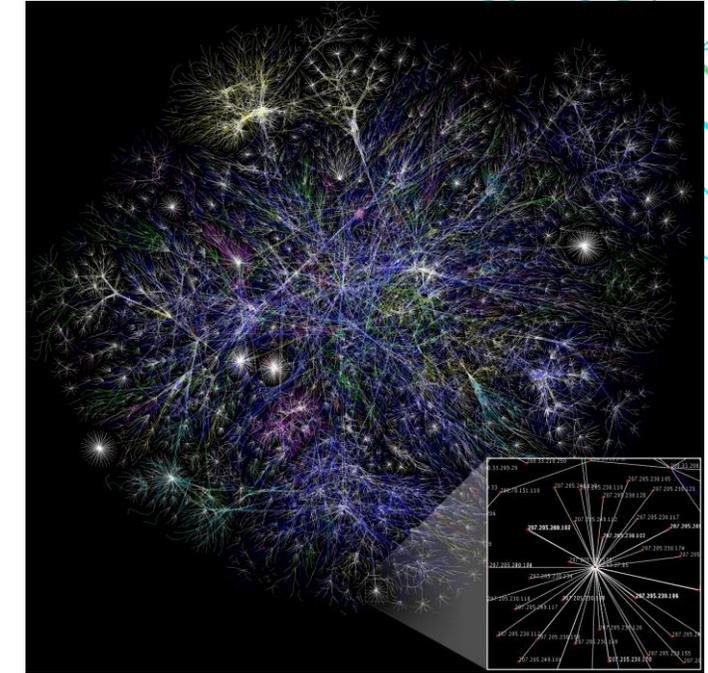
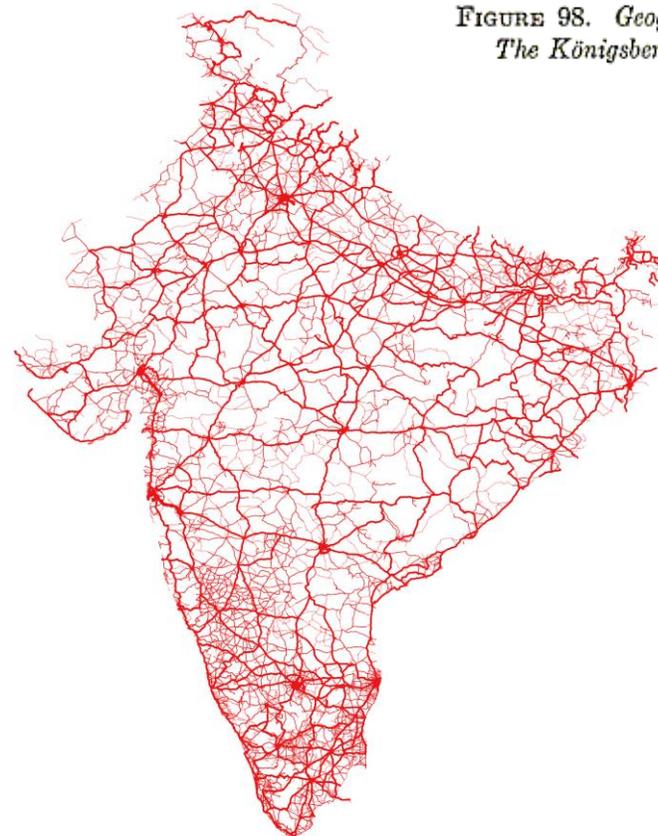


FIGURE 98. Geographic Map:
The Königsberg Bridges.



Scenario 1: grafo del web

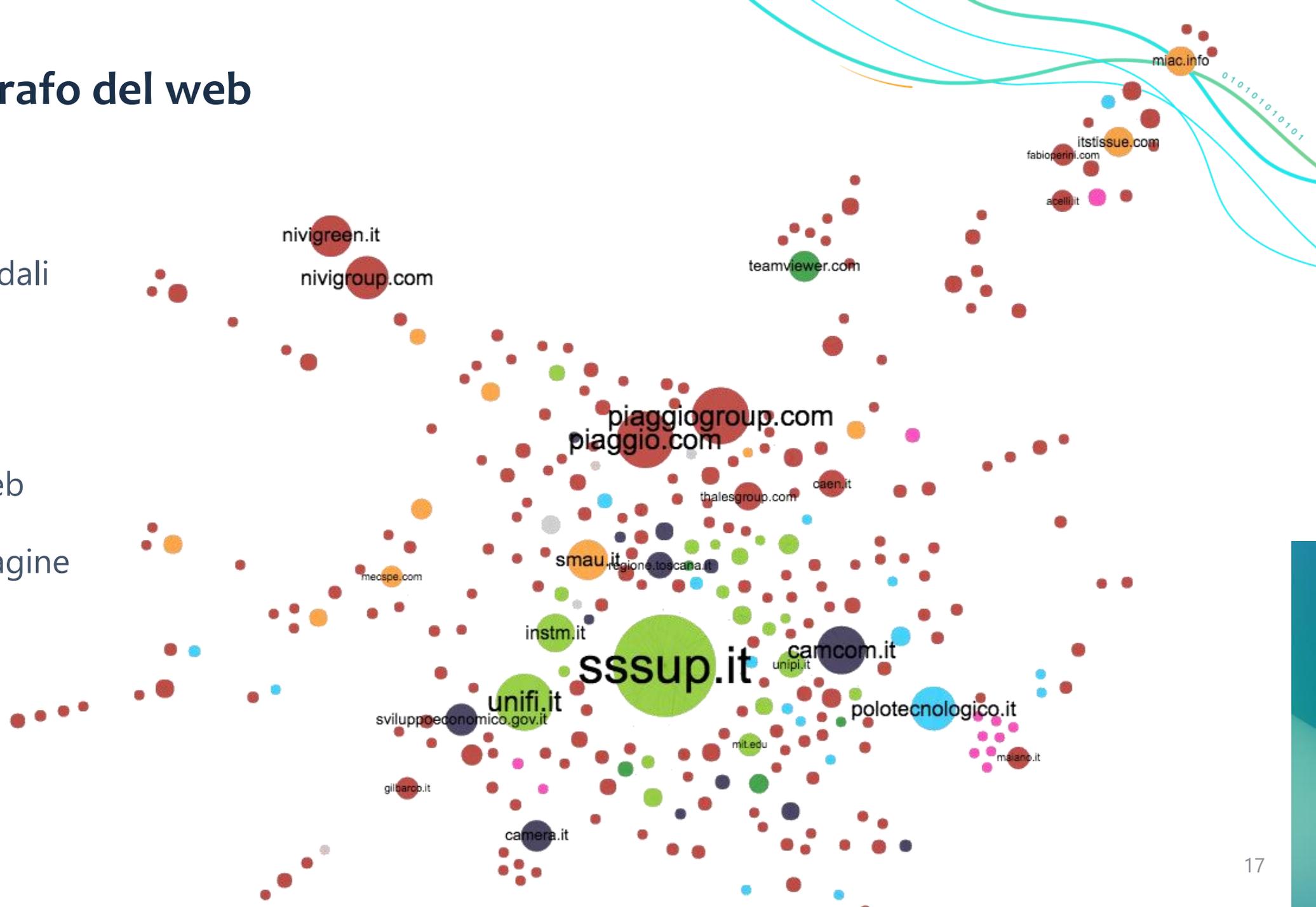
1B pagine web

6M domini

1M domini aziendali

Nodi: pagine web

Archi: link tra pagine



Scenario 2: grafo semantico tra pagine Wikipedia

The **Mona Lisa** is a **sixteenth century** **oil painting** created by **Leonardo**. It's held at the **Louvre** in **Paris**.

0 persons

1 work

0 organisations

1 place

0 events

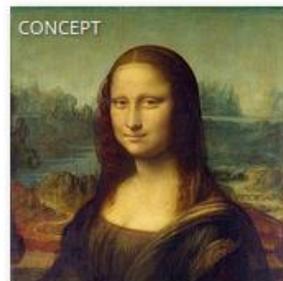
4 concepts



Mona Lisa



16th century



Oil painting



Leonardo da Vinci



Louvre



Paris

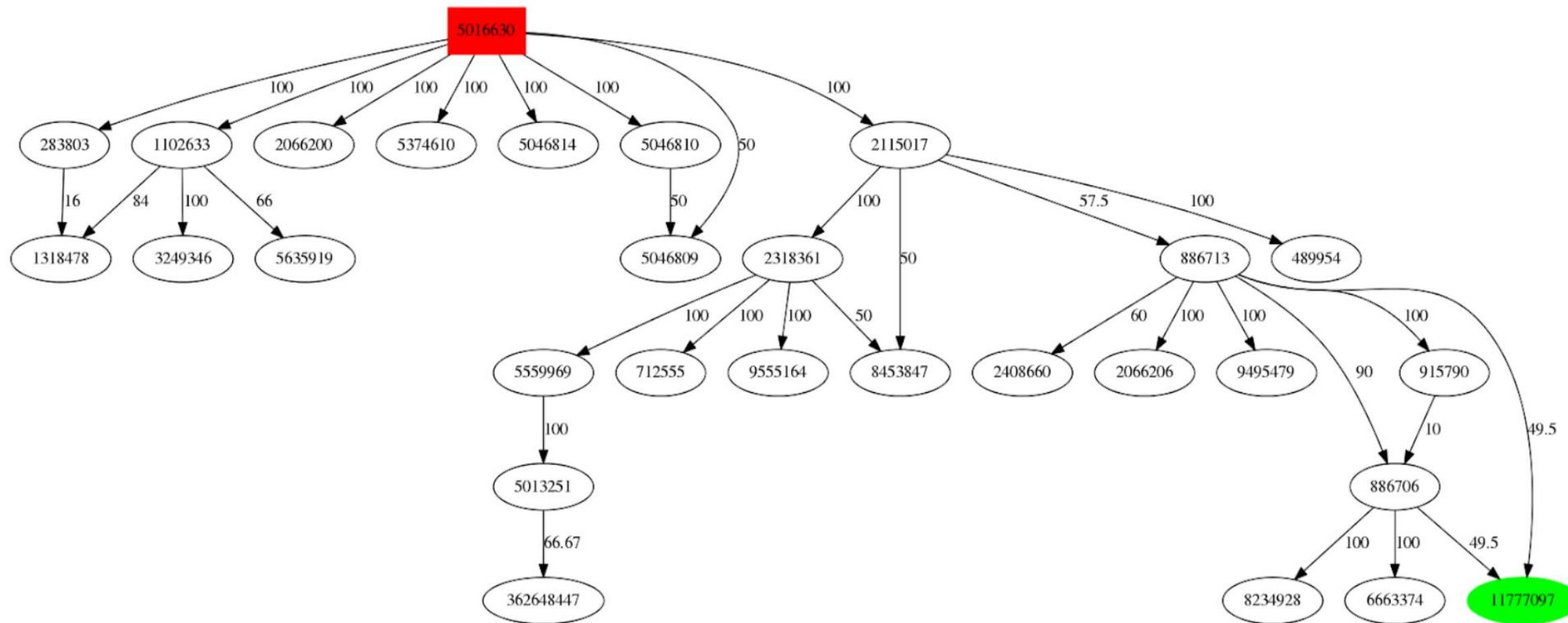
Caso d'uso: analisi di testi

Nodi: pagine Wikipedia

Archi: relatedness tra pagine

Scenario 3: grafo di controllo societario

Chi controlla una società? Chi ne controlla quote > 50%



Nodi controllati dal nodo 5016630

Dati territoriali Cerved

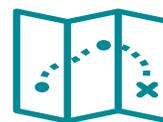
Dati su imprese, immobili e persone fisiche, raggruppati a livello di cella censuaria, per qualificare e descrivere il territorio italiano con score standard e personalizzati.



Caratteristiche
popolazione



Caratteristiche
imprese

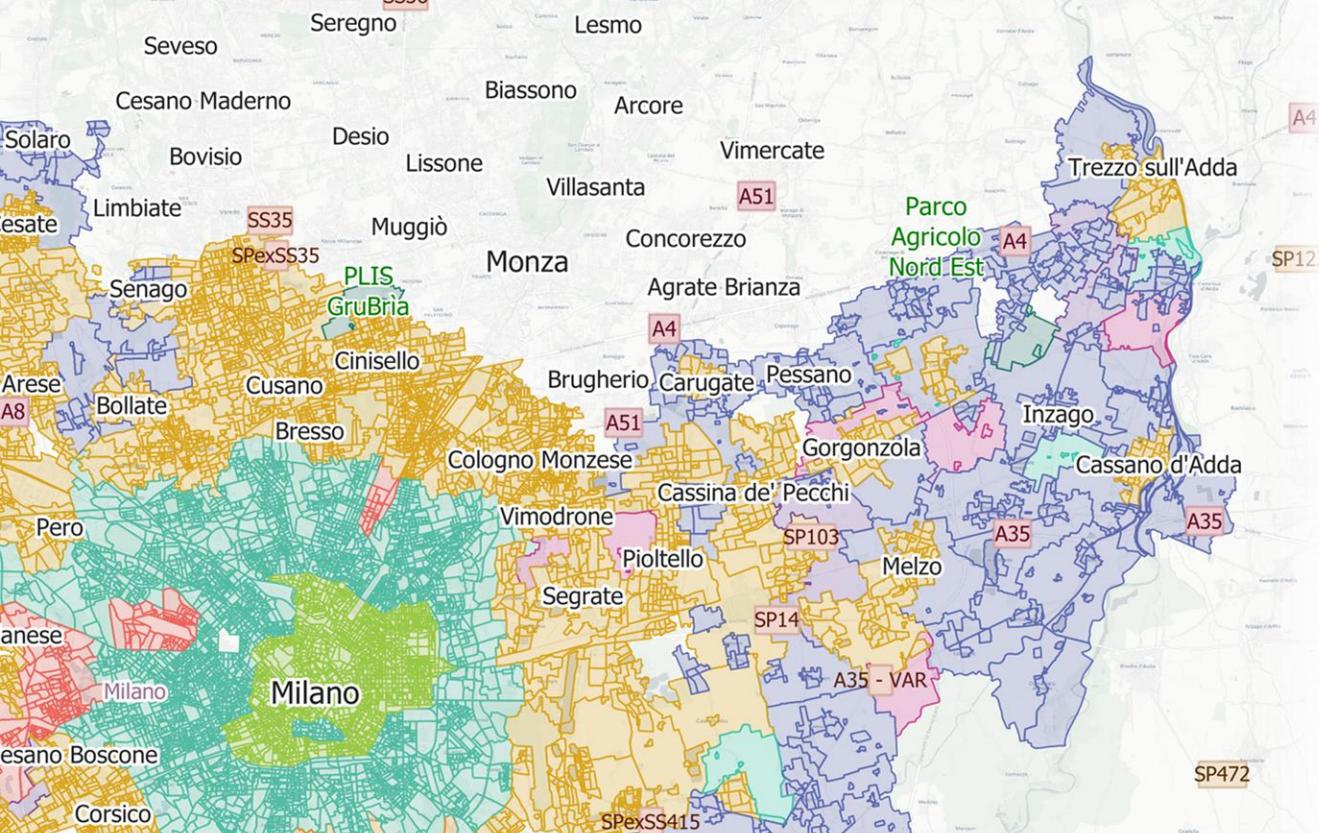


Caratteristiche
territorio



Mercato
immobiliare





Titoli dei Segmenti

- Away From It All
- Blue Collars Living in Nice Outskirts
- City Dwellers and Young Urbanites
- City Insiders
- Comfortable Retirement
- Inner and Isolated Hardship
- Outer Suburb Condo
- Rural Farming Idilly
- Rural Multigenerational Successful Families
- Suburban Mixed Middle Class

Clustering geografico

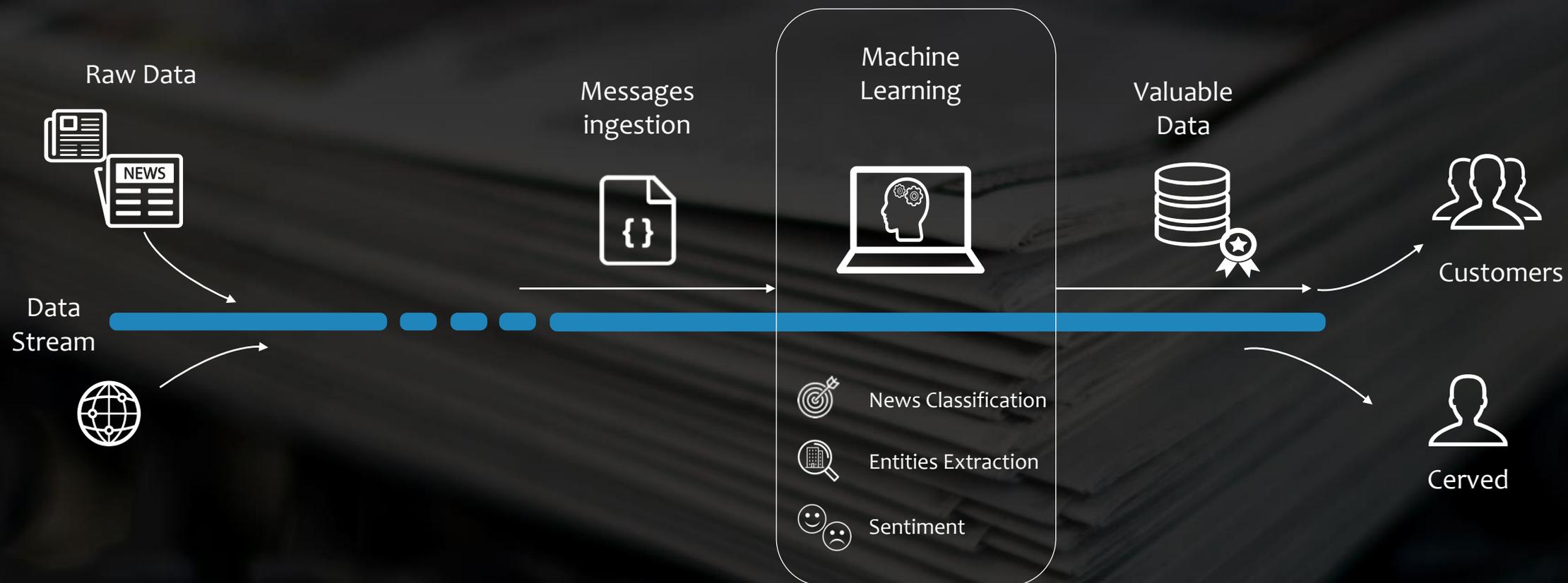
1. Classificazione delle sezioni in base a un processo di segmentazione geodemografica.
2. Identificare zone del territorio con particolari concentrazioni di attività economiche associate ad una stessa filiera produttiva.

Problema:

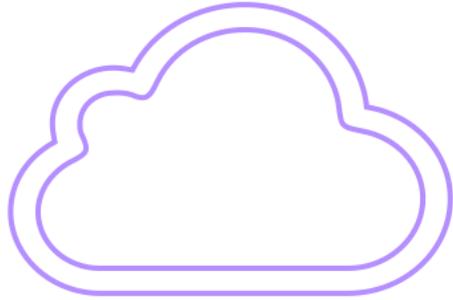
- Definizione dei parametri di ricerca.
- Scelta delle trasformazioni sui dati.
- Scelta di metriche di validazione.

Machine Learning per News Online

Estrarre informazioni di valore per il Business in tempo reale dalle notizie online.

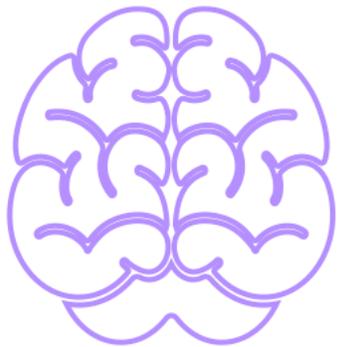


News - Sfide da affrontare



Tecnologiche

- Grossi volumi di dati da processare in near real-time.
- Scalabilità della soluzione di fronte a nuove sorgenti e richieste.
- Utilizzo delle ultime tecnologie Cloud.



Algoritmiche

- Modelli in grado di lavorare su testi lunghi e in lingua italiana.
- Contesto di weak learning per molti task.
- Utilizzo di modelli di Deep Learning evoluti.



Troppe cose da raccontare??

Se siete interessati fateci le domande e scriveteci! :)

Michele Voto: michele.voto@cerved.com

Roberto Santoro: roberto.santoro@cerved.com

Gabriele Sotto: gabriele.sotto@cerved.com



Thank you

