



**CONVEGNO  
EX MACHINA  
L'INTELLIGENZA ARTIFICIALE  
TRA AZIONE AMMINISTRATIVA  
E GIUSTIZIA 4.0**

Piattaforme Analitiche nell'emergente  
campo delle Scienze Giuridiche  
Computazionali

Alfredo Pulvirenti  
Università degli studi di Catania

# Chi sono

- Docente di Informatica dell'Università di Catania



- **Campi di ricerca**

- Analisi dei dati, Big Data e Data Mining
- Sistemi di raccomandazione
- Network mining
- Bioinformatica



<http://www.dmi.unict.it/~apulvirenti/site/>



Antudo  
Cloud  
Software



## Sommario

- Background
- Machine learning
- Applicazioni e prospettive nel contesto giuridico:
  - Possiamo migliorare le decisioni dei giudici?
  - Possiamo aiutarli nel prendere la loro decisione?

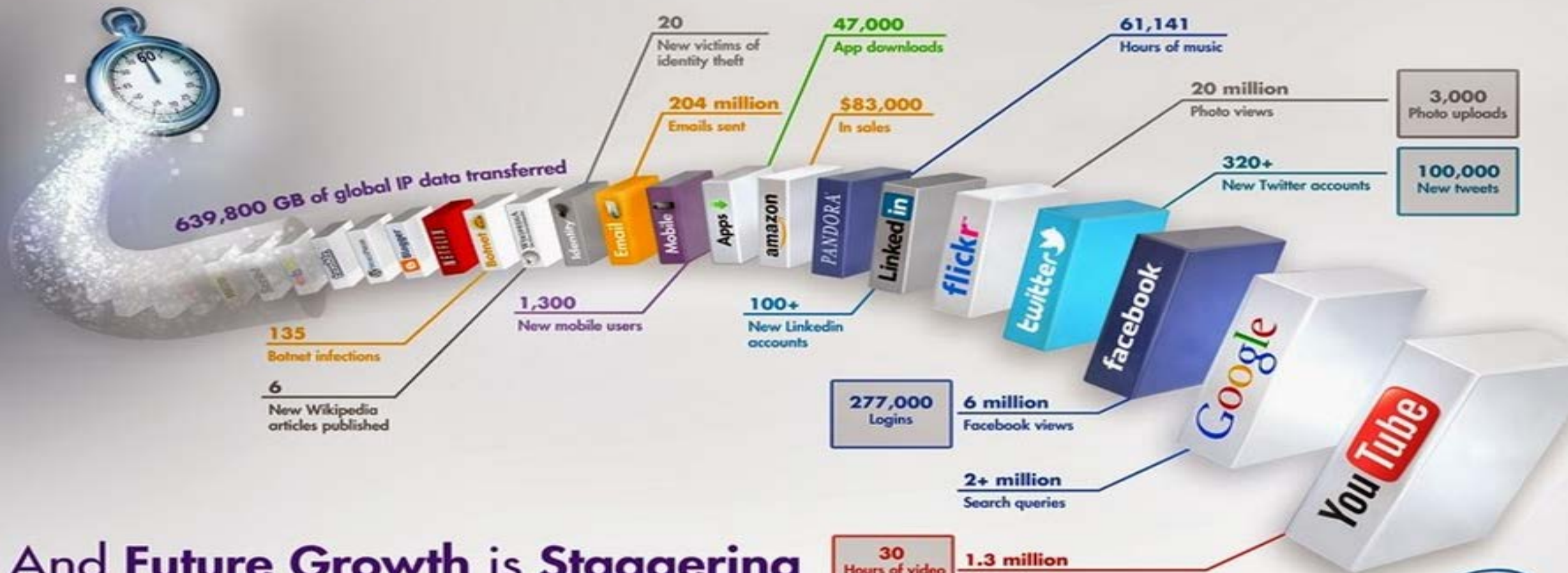
# L'era dei dati

- Prodotti velocemente
- Eterogenei
- Potenzialmente ricchi di informazioni

**Big data:** Qualsiasi processo di dati in cui la dimensione dei dati stessi è un problema: conservare, trasmettere, elaborare su scala.



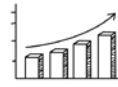
# What Happens in an Internet Minute?



## And Future Growth is Staggering



# Applicazioni



**Business:** pubblicità, modellizzazione dei clienti, CRM, e-Commerce, rilevamento delle frodi, investimenti, produzione, sport / intrattenimento, telecomunicazioni (telefono e comunicazioni), marketing mirato, ecc.



**Scienza:** astronomia, bioinformatica, scoperta di farmaci, ecc.



**Health:** riposizionamento farmaci, valutazione del trattamento, modificazione del comportamento, ecc.



**Governo:** controllo sociale, forze dell'ordine, imposizione fiscale, governo aperto, ecc.



**Web:** motori di ricerca, bots, ecc.

# Predictive Analytics



Tecnologia che **impara dall'esperienza** (dati) per predire il comportamento futuro degli individui al fine di guidare decisioni migliori

*Da Predictive Analytics:  
The Power to Predict Who Will Click, Buy, Lie, or Die  
di Eric Siegel*





Termini  
correlati

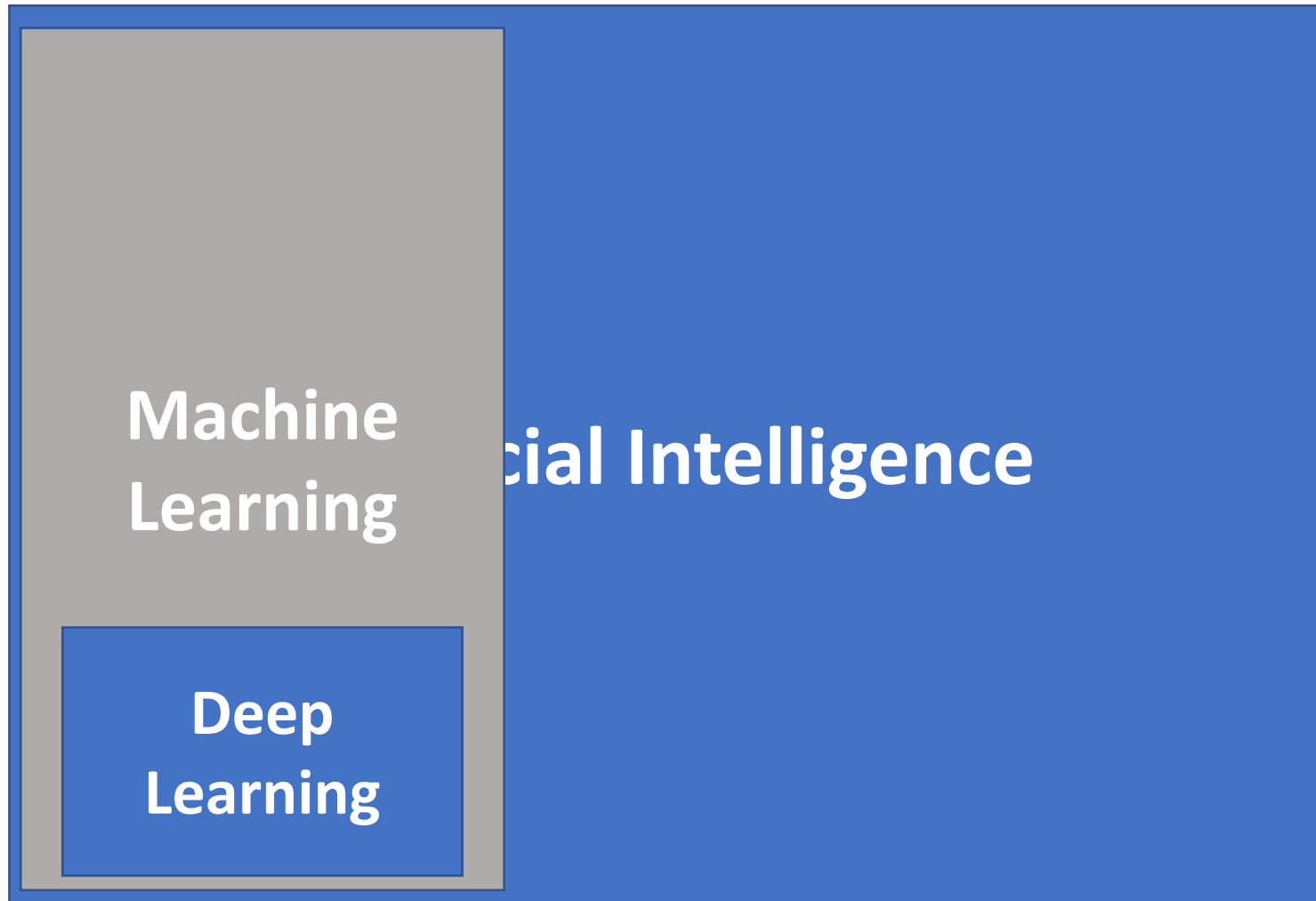
Data mining

Statistica

Data science

Business intelligence

# Capacità tecniche emergenti



# Termini correlati



data  
iku

## AN INTRODUCTION TO KEY DATA SCIENCE CONCEPTS

Definitions of 10 fundamental terms for data science and machine learning.



### *model (n)*

[ˈmɒdəl] / noun

1. a mathematical representation of a real world process; a predictive model forecasts a future outcome based on past behaviors.



### *algorithm (n)*

[ˈælgəˌrɪðəm] / noun

1. a set of rules used to make a calculation or solve a problem.



### *training (v)*

[ˈtreɪnɪŋ] / verb

1. the process of creating a model from the training data. The data is fed into the training algorithm, which learns a representation for the problem, and produces a model. Also called "learning".



### *regression (n)*

[rəˈɡreʃən] / noun

1. a prediction method whose output is a real number, that is, a value that represents a quantity along a line. Example: predicting the temperature of an engine or the revenue of a company.



### *classification (n)*

[klæsəfəˈkeɪʃən] / noun

1. a prediction method that assigns each data point to a predefined category, e.g., a type of operating system.



### *target (n)*

[ˈtɑːɡət] / noun

1. in statistics, it is called the dependent variable; it is the output of the model or the variable you wish to predict.



### *training set (n)*

[ˈtreɪnɪŋ set] / noun

1. a dataset used to find potentially predictive relationships that will be used to create a model.



### *test set (n)*

[test set] / noun

1. a dataset, separate from the training set but with the same structure, used to measure and benchmark the performance of various models.



### *feature (n)*

[ˈfi:tʃər] / noun

1. also known as an independent variable or a predictor variable, a feature is an observable quantity, recorded and used by a prediction model. You can also engineer features by combining them or adding new information to them.



### *overfitting (v)*

[ˌoʊvərˈfi:tɪŋ] / verb

1. a situation in which a model that is too complex for the data has been trained to predict the target. This leads to an overly specialized model, which makes predictions that do not reflect the reality of the underlying relationship between the features and target.



# Machine Learning



- Trovare **pattern** nei dati
- Costruire dei **modelli** basati sui pattern
- **Classificare** i dati nuovi usando i pattern

# Learning quando acquistiamo on-line

Informatica › Monitor



## NEC X981UHD Monitor

di NEC

★★★★☆ 120 recensioni

**Attualmente non disponibile**

Ancora non sappiamo quando l'articolo sarà disponibile.

- NUOVO Nec 60003721 MULTIDISPLAY 500CD/M 24/7 PROC

[Visualizza altri dettagli prodotto](#)

[Confronta con articoli simili](#)

[Segnala informazioni inesatte](#)

**Informazioni sulla spedizione:** P voluminosi venduti su Amazon.it articoli potrebbero essere soggetti a tasse di spedizione. [Visualizza dettagli.](#)

★★★★☆ nek

Da lenzo il 25 settembre 2015

Ho acquistato questo articolo pensando fosse l'ultimo album di nek (il cantante), appena ho visto il camioncino sotto casa e considerando il prezzo ho pensato che montassero un palco e cantasse nek in persona, delusione totale. Ora lo sto usando come parete tv per il mio 50" al plasma hinnohit

## La recensione più critica

[Visualizza tutte le 16 recensioni critiche](#)

91 persone l'hanno trovato utile

★★★★☆ **Acquistato per errore**

Da marco il 4 giugno 2016

L'ho acquistato per errore a un Sunday Bloody Sunday di Amazon, insieme a un cavetto HDMI AmazonBasics, dannata frenesia. Poi a mente fredda ho deciso di confermare l'acquisto. Crepi l'avarizia, mi sono detto, sarò pure genovese, sarò pure una sottomarca, ma potrò ben buttare via trecento euro per un nuovo TV, dopo una vita di sacrifici.

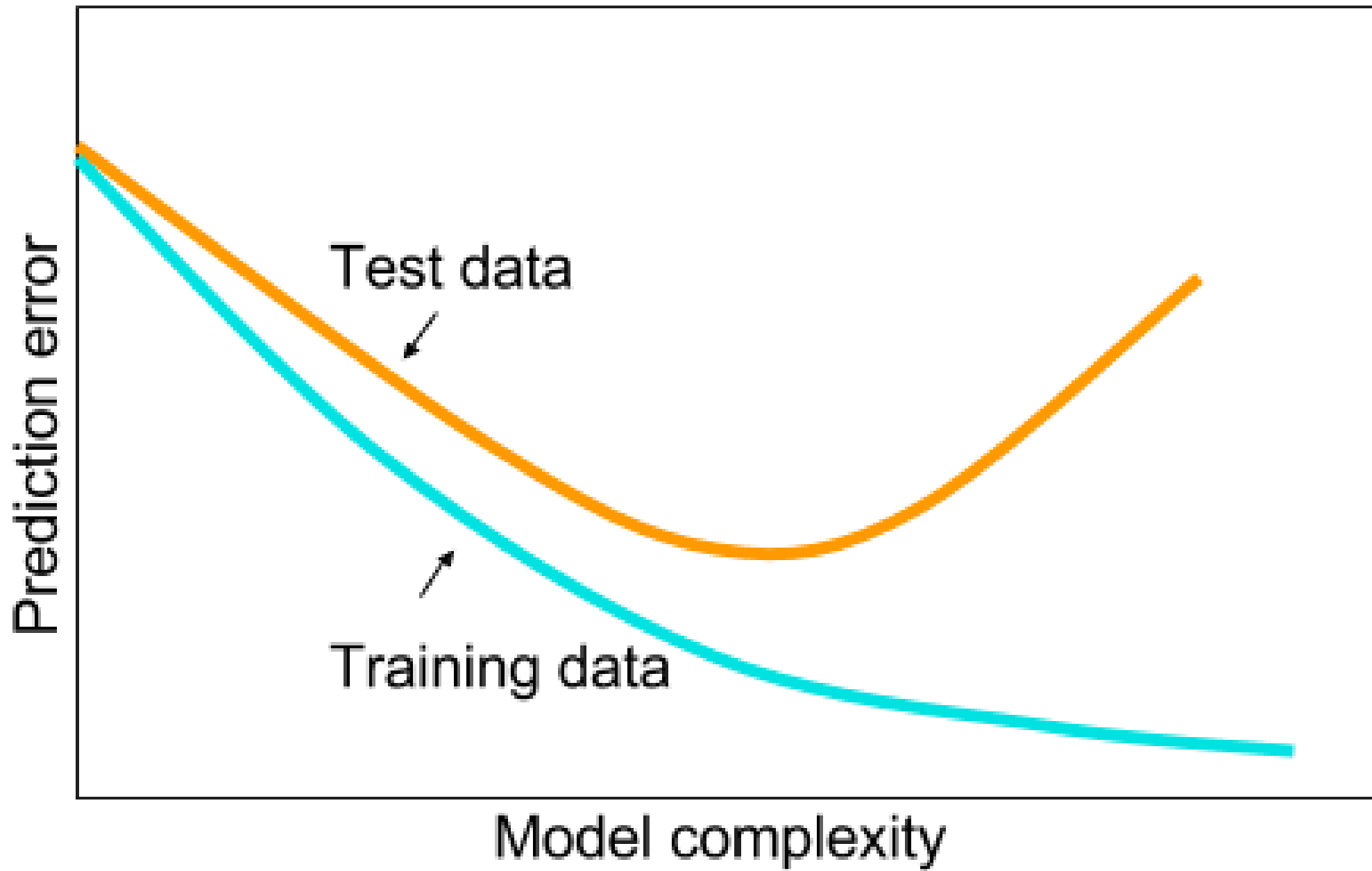
Che qualcosa fosse andato storto l'ho capito per il sopralluogo della Guardia di Finanza, avvenuto qualche minuto dopo.

La mia vita è cambiata, potete immaginarlo.

# Raccomandazione contenuti (es. Netflix)



Joe	1	1	1	0	0
Jim	0	3	3	0	0
John	4	4	4	0	0
Jack	5	5	0	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	0
Jane	0	0	0	2	2



# Cosa c'è di nuovo: i media sono ora testo





# Le immagini oggi valgono 1K di JSON

The screenshot shows the Google Cloud documentation page for the Cloud Vision API. The page title is "Drag and drop" and it includes a "Try the API" section with a dashed box for uploading an image. The page also features a sidebar with navigation links, a search bar, and a feedback section.

**Google Cloud** Search


Why Google Products Solutions Launcher Pricing Security Customers **Documentation** Support Partners

Cloud Vision API > Documentation ☆☆☆☆☆ [SEND FEEDBACK](#)

## Drag and drop

Use the application below to return image annotations for your image file. Click the **JSON** tab to view the raw response.

- Maximum file size is 4MB.
- Your browser must have JavaScript enabled.

**Try the API** Drag image file here or Browse from your computer 


Was this page helpful? Let us know how we did: ☆☆☆☆☆ [SEND FEEDBACK](#)


Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 3.0 License](#), and code samples are licensed under the [Apache 2.0 License](#). For details, see our [Site Policies](#). Java is a registered trademark of Oracle and/or its affiliates.

Last updated February 12, 2018.

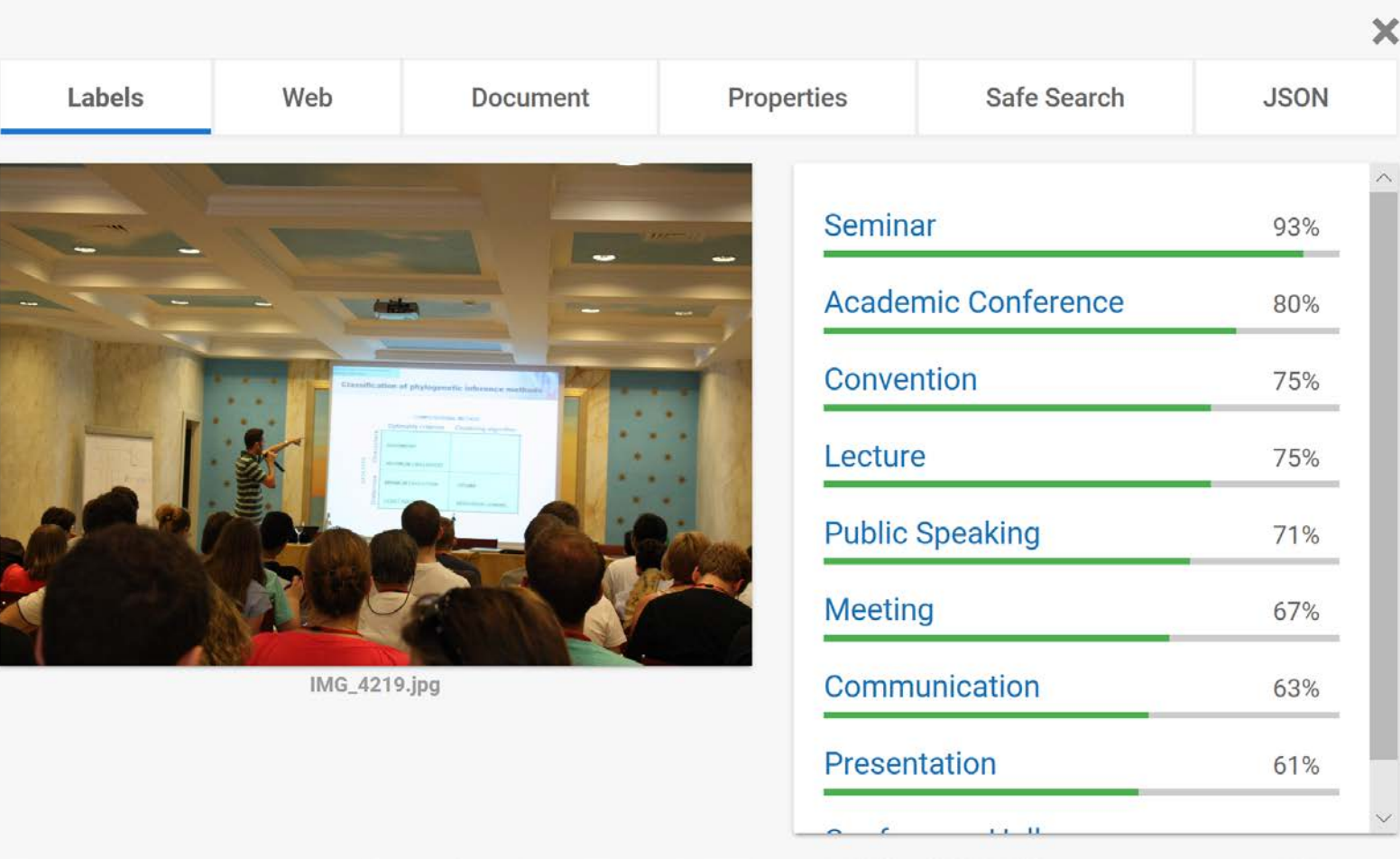
**Cloud Vision API**  
Product Overview  
Documentation

**Quickstarts**  
All Quickstarts  
Quickstart: Embedded API Explorer  
[Quickstart: Drag and Drop](#)

**How-to Guides**  
All How-to Guides  
Enable the API  
Authenticate  
Make a Request  
Detect Text (OCR)  
Detect Labels  
Detect Other Features  
Beta Features   
Base64 Encode

**APIs & Reference**  
All APIs & References  
Client Libraries  
Migrating to Python Client Library v0.25.1 

# Le immagini oggi valgono 1K di JSON

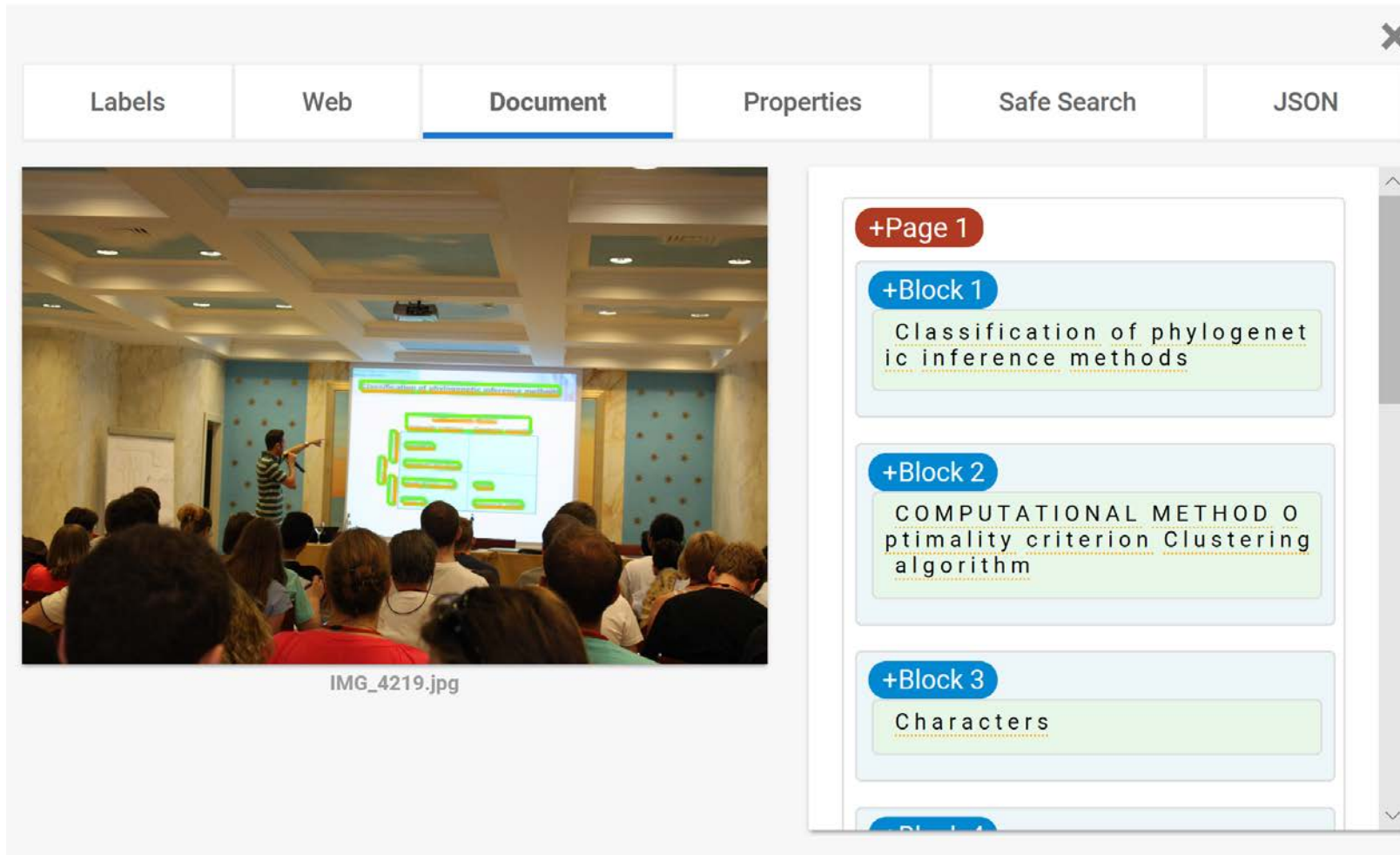


The screenshot displays the Google Cloud Vision API interface. At the top, there are tabs for 'Labels', 'Web', 'Document', 'Properties', 'Safe Search', and 'JSON'. The 'Labels' tab is selected. Below the tabs is a large image of a seminar in progress, with a presenter at the front and an audience. The image is labeled 'IMG\_4219.jpg'. To the right of the image is a list of classification labels with their corresponding confidence percentages, each accompanied by a green progress bar.

Label	Confidence
Seminar	93%
Academic Conference	80%
Convention	75%
Lecture	75%
Public Speaking	71%
Meeting	67%
Communication	63%
Presentation	61%

<https://cloud.google.com/vision/docs/drag-and-drop>

# Le immagini oggi valgono 1K di JSON

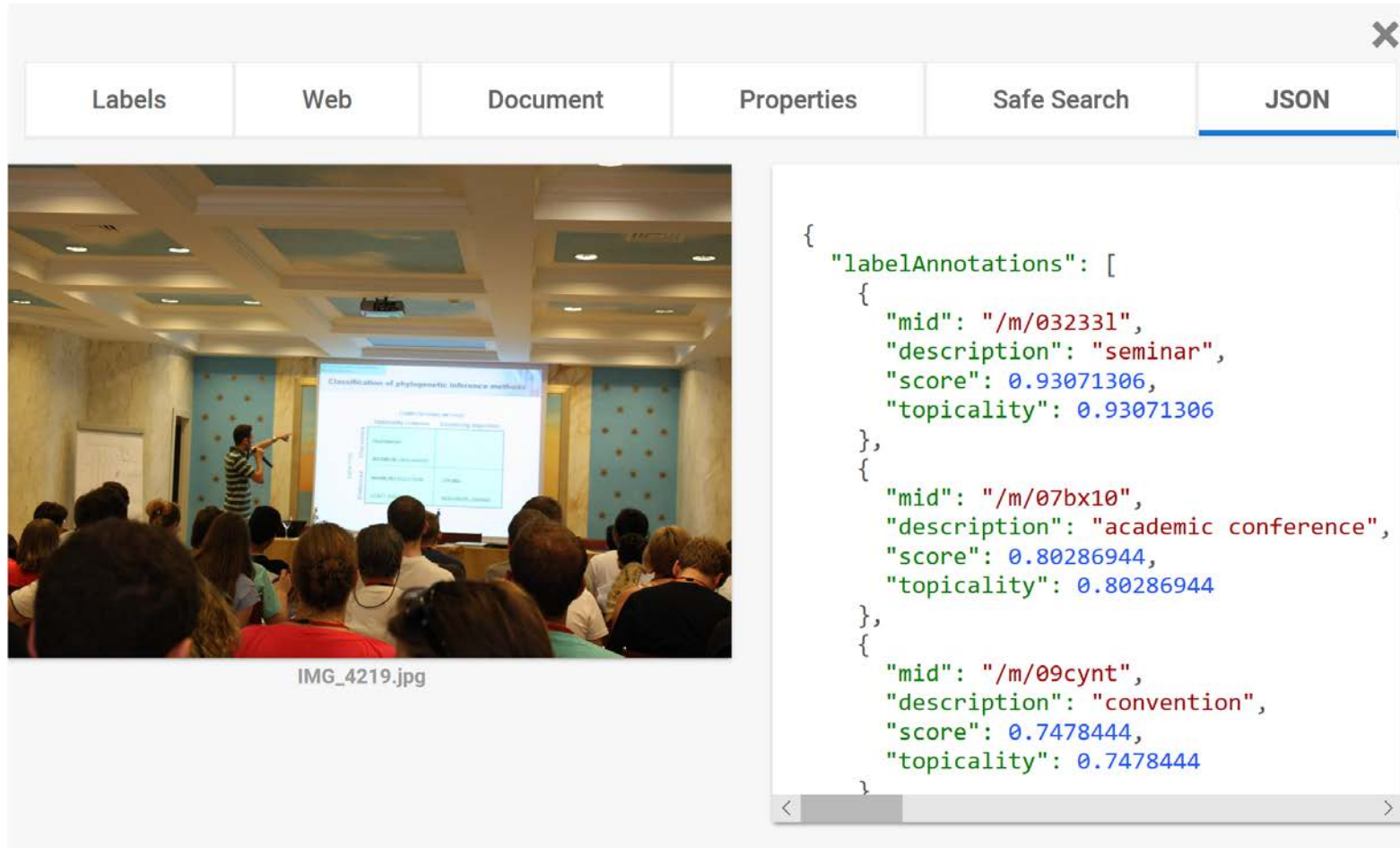


The screenshot displays the Google Cloud Vision API interface. At the top, there are tabs for 'Labels', 'Web', 'Document', 'Properties', 'Safe Search', and 'JSON'. The 'Document' tab is selected. Below the tabs, a photograph of a lecture hall is shown, with a presenter at the front and an audience. The image is labeled 'IMG\_4219.jpg'. To the right of the image, a list of detected text blocks is displayed under the heading '+Page 1'. The blocks are:

- +Block 1: Classification of phylogenetic inference methods
- +Block 2: COMPUTATIONAL METHOD Optimality criterion Clustering algorithm
- +Block 3: Characters

<https://cloud.google.com/vision/docs/drag-and-drop>

# Le immagini oggi valgono 1K di JSON



The screenshot displays the Google Cloud Vision API interface. At the top, there are tabs for 'Labels', 'Web', 'Document', 'Properties', 'Safe Search', and 'JSON'. The 'JSON' tab is selected. Below the tabs, there is a preview of an image showing a seminar in progress. The image is labeled 'IMG\_4219.jpg'. To the right of the image, the JSON output is displayed, showing three labels with their respective scores and topicalities.

```
{
  "labelAnnotations": [
    {
      "mid": "/m/032331",
      "description": "seminar",
      "score": 0.93071306,
      "topicality": 0.93071306
    },
    {
      "mid": "/m/07bx10",
      "description": "academic conference",
      "score": 0.80286944,
      "topicality": 0.80286944
    },
    {
      "mid": "/m/09cynt",
      "description": "convention",
      "score": 0.7478444,
      "topicality": 0.7478444
    }
  ]
}
```


<https://cloud.google.com/vision/docs/drag-and-drop>

# Guida Visuale al Machine Learning

<http://www.r2d3.us/una-introduzione-visuale-al-machine-learning-1/>

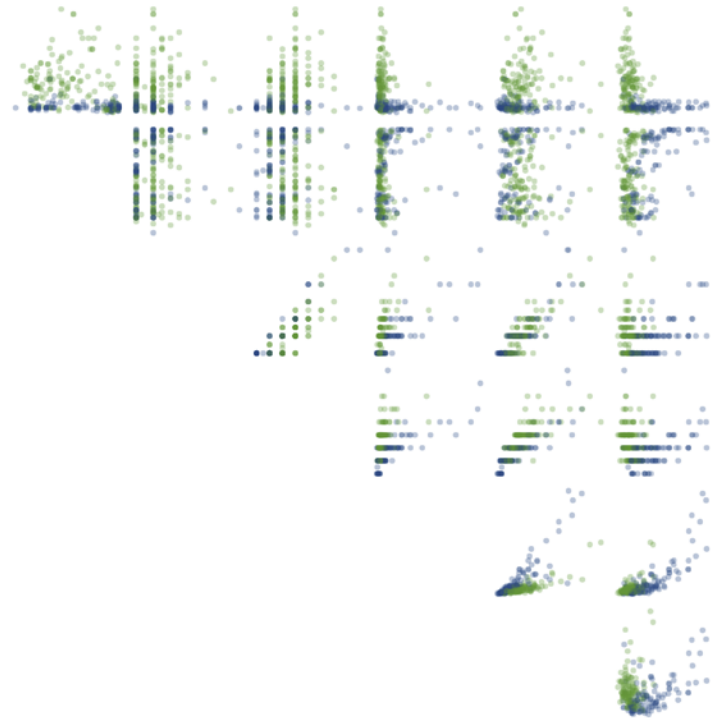
R2  
D3

## Una introduzione visuale al machine learning

 Italiano

In apprendimento automatico (*machine learning* in Inglese), i computer applicano tecniche di **apprendimento statistico** per identificare automaticamente delle strutture nei dati. Queste tecniche possono essere usate per effettuare predizioni con elevata accuratezza.

*Continua a scorrere.* Usando dei dati che riportano informazioni sulle case, creeremo un

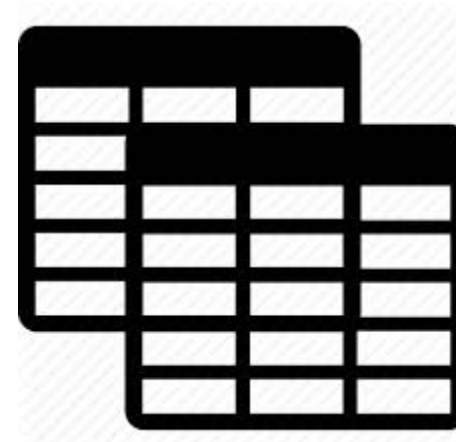
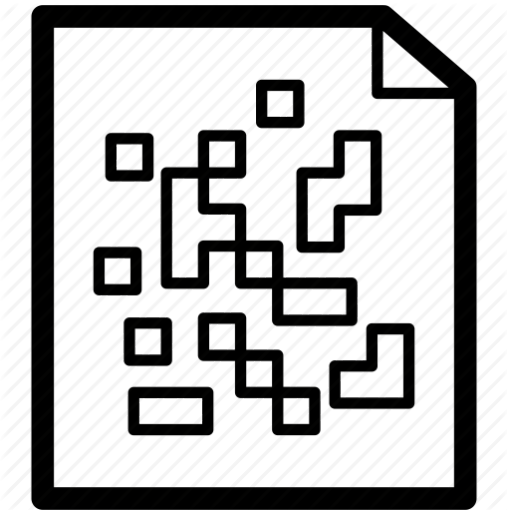


# Ma...

- Apprendimento automatico sorprendentemente efficace in un'ampia varietà di compiti tradizionalmente associati all'intelligenza umana
- Se questi strumenti possono essere applicati con successo a compiti basilari che implicano la visione e il linguaggio umano, **perché non prendere anche decisioni umane più complesse?**

# Capacità tecniche emergenti

Passare dal testo a dati non strutturati arricchiti



# Un modello predittivo per il supporto alle decisioni giudiziarie /1

- Studio sistematico **previsione esito casi giudicati** dalla Corte Europea dei diritti umani basati esclusivamente sul contenuto testuale
- Formulazione **classificazione binaria**: attraverso l'analisi del contenuto testuale estratto da un caso si determina automaticamente se c'è stata una violazione di un articolo della convenzione sui diritti umani



# Da dove si parte?

- **Procedura.** Procedura seguita dalla Corte, dalla presentazione della domanda individuale fino alla pronuncia della sentenza
  - **I fatti:** Le circostanze del caso. Leggi (nazionali) pertinenti
  - **Le leggi:** La sezione di legge è focalizzata sulla considerazione dei meriti del caso, attraverso l'uso di argomentazioni legali
- 
- **Disposizioni operative:** La corte annuncia l'esito del caso: decisione in base alla quale **ha avuto** o **non ha avuto** luogo una violazione di alcuni articoli della Convenzione

# I dati

Dataset contenente i casi relativi agli articoli 3, 6 e 8 della Convenzione

**Table 1** Articles of the Convention and number of cases in the data set. Article numbers, Convention right that each article protects and the number of cases in our data set.

Article	Human Right	Cases
3	Prohibits torture and inhuman and degrading treatment	250
6	Protects the right to a fair trial	80
8	Provides a right to respect for one's "private and family life, his home and his correspondence"	254

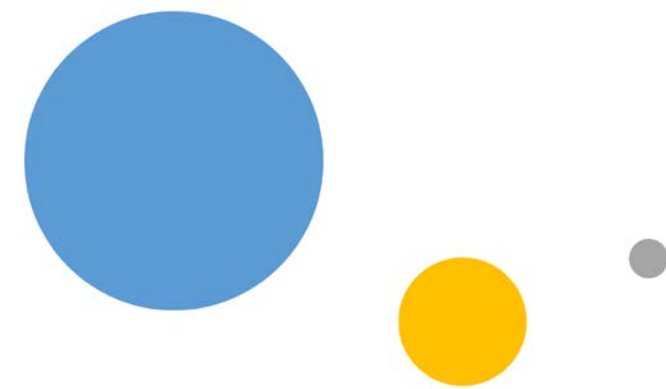


**Table 2 Accuracy of the different feature types across articles.** Accuracy of predicting violation/non-violation of cases across articles on 10-fold cross-validation using an SVM with linear kernel. Parentheses contain the standard deviation from the mean. Accuracy of random guess is .50. **Bold** font denotes best accuracy in a particular Article or on Average across Articles.

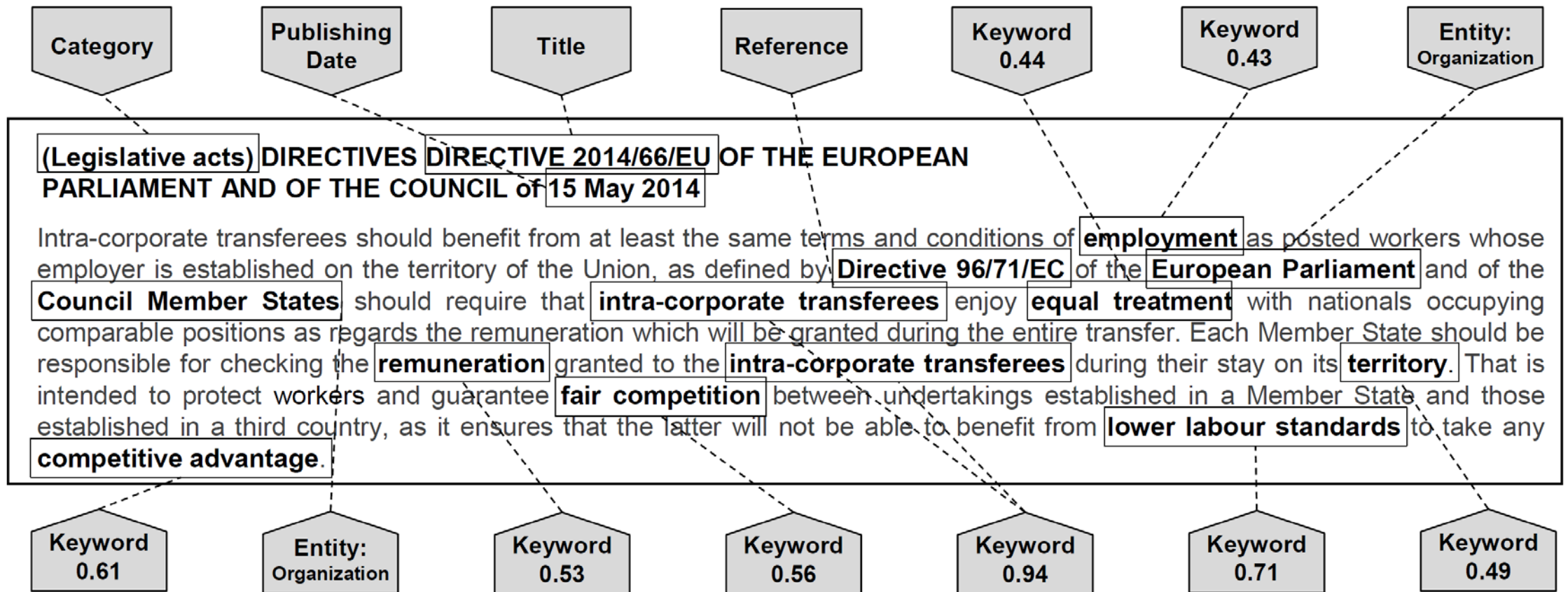
Feature Type		Article 3	Article 6	Article 8	Average
N-grams	Full	.70 (.10)	.82 (.11)	.72 (.05)	.75
	Procedure	.67 (.09)	.81 (.13)	.71 (.06)	.73
	Circumstances	.68 (.07)	.82 (.14)	.77 (.08)	.76
	Relevant law	.68 (.13)	.78 (.08)	.72 (.11)	.73
	Facts	.70 (.09)	.80 (.14)	.68 (.10)	.73
	Law	.56 (.09)	.68 (.15)	.62 (.05)	.62
Topics		<b>.78 (.09)</b>	.81 (.12)	.76 (.09)	.78
Topics and circumstances		.75 (.10)	<b>.84 (0.11)</b>	<b>.78 (0.06)</b>	<b>.79</b>

# Considerazioni



- Risultati promettenti
  - Utili come **strumento di assistenza rapida** per identificare i casi ed estrarre schemi che portano a determinate decisioni
- 

# Ma possiamo andare oltre: associare una semantica



# Come?

## Sistemi di annotazione

Il Presidente Mattarella passa al vaglio il decreto Salvini appena deliberato, all'unanimità, dal governo di Giuseppe Conte.

Many links



Few links

Reset

**TAGME!**

Tagged text

Topics

Il [Presidente Mattarella](#) passa al vaglio il [decreto Salvini](#) appena deliberato, all'unanimità, dal [governo](#) di [Giuseppe Conte](#).

**Governo della Repubblica**

Il Governo della Repubblica Italiana è l'organo complesso del sistema di governo, composto dal Presidente della Repubblica, dal Consiglio dei Ministri, dai ministri e dai Ministri, che insieme...

**Sergio Mattarella**

Dal 1983 al 2008 è stato deputato, prima per la Democrazia Cristiana (di cui fu vicesegretario) e poi per il Partito Popolare Italiano, La Margherita e il Partito Democratico. Ha ricoperto la carica di...

# Un modello predittivo per il supporto alle decisioni giudiziarie /2



- Ogni anno, negli Stati Uniti, la polizia **arresta oltre 10 milioni di persone** (FBI, 2016)
- Dopo l'arresto, i **giudici decidono** se gli imputati devono aspettare in prigione mentre viene deciso il loro destino legale
- I casi possono richiedere diversi mesi in media per risolverli, questa è una decisione che ha importanti **conseguenze** sia per gli **imputati** che per la **società** nel suo complesso



# Un modello predittivo per il supporto alle decisioni giudiziarie /2



I giudici dovrebbero, per legge, basare la loro decisione unicamente su una previsione: cosa farà l'imputato se rilasciato? Fuggirà? Commetterà un altro crimine?



Altri fattori, come se il convenuto sia colpevole, non dovrebbero influenzare questa decisione

# Ancora una volta il Machine Learning

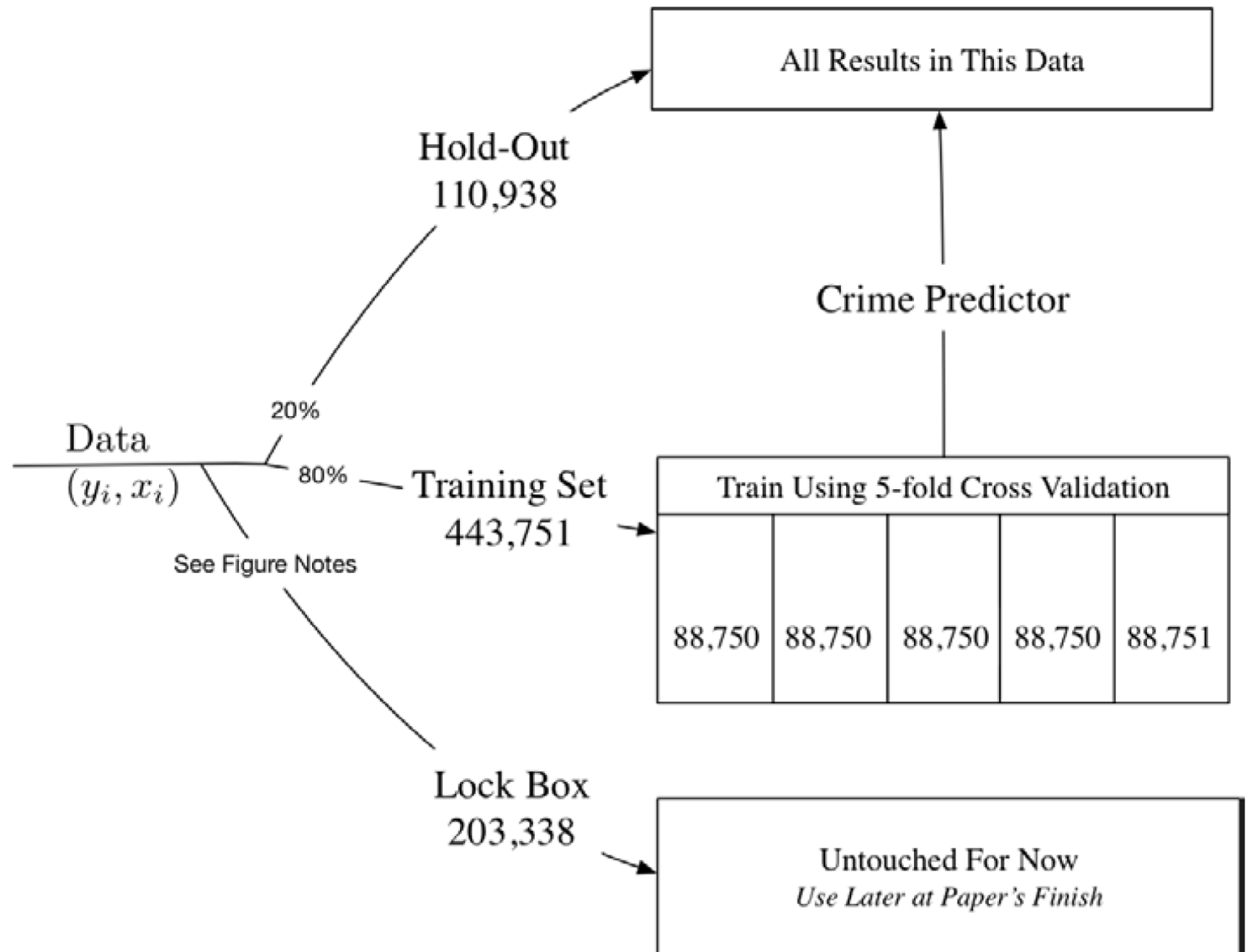


- **Associare decisione cauzione ad un pronostico** (commissione di un reato) applicazione ideale per il machine learning
- Si può definire un **problema di predizione** dove si introduce una funzione che calcola un rischio e lo associa all'imputato
- Algoritmo di apprendimento automatico **addestrato su alcune caratteristiche degli imputati** per prevedere il rischio di crimine

# I dati

- Set di **758.027 imputati arrestati** a New York City tra il 2008 e il 2013
- **Informazioni dettagliate su imputati**: se stati rilasciati prima del processo e, in caso affermativo, se hanno continuato a commettere reati dopo il rilascio
- In questo caso, gli input (l'analogo della foto in un'applicazione di riconoscimento facciale) sono gli **attributi del convenuto a disposizione dei giudici** quando decidono

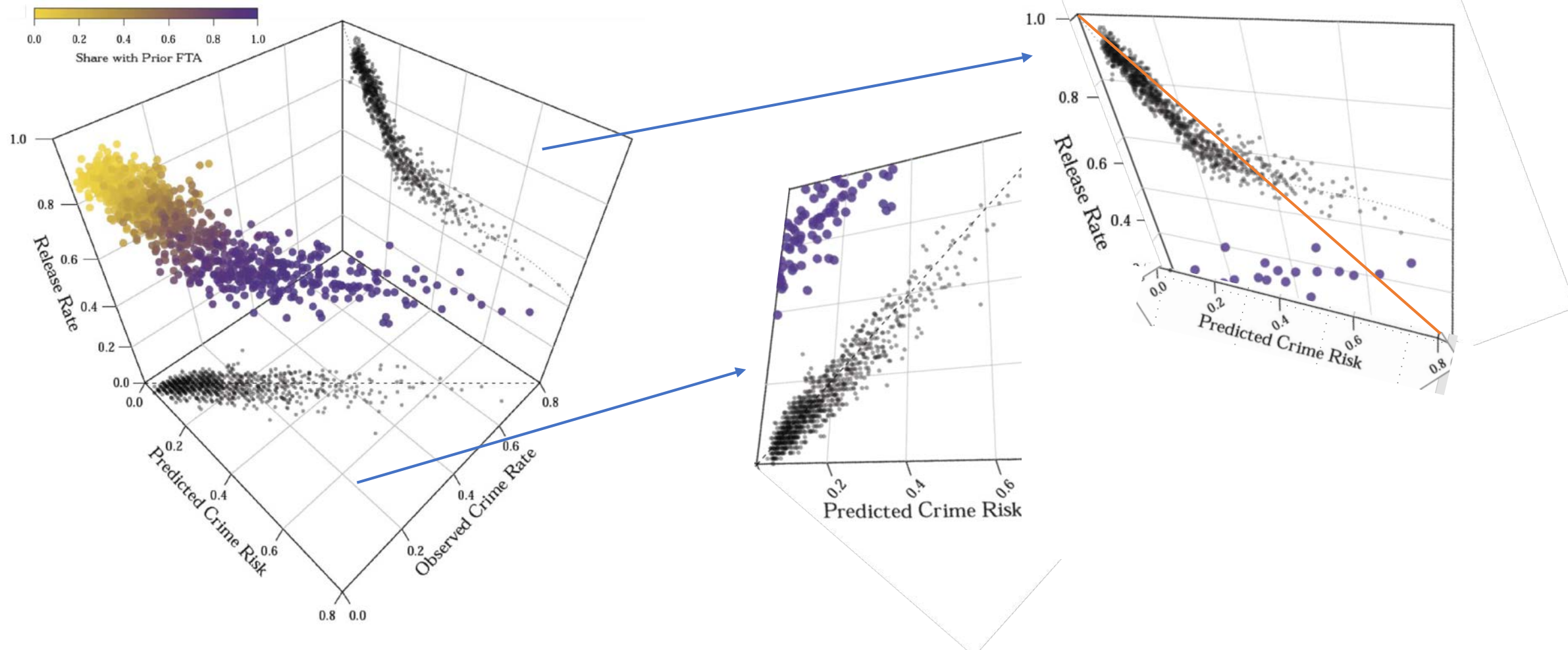
# Train/Test



# Cosa possiamo valutare?


- Queste previsioni possono aiutarci a capire come **migliorare le decisioni dei giudici**
- Bias: mentre possiamo fare previsioni per tutti gli imputati, **abbiamo solo esiti per gli imputati liberati** (il giudice selettivamente nasconde delle variabili)

# Decisioni dei giudici vs predizioni della macchina



# Esiste una differenza significativa tra la decisione del giudice e quella del modello di machine learning?

Le decisioni del giudice spesso non sono conformi alle previsioni dell'algoritmo



Molti degli imputati contrassegnati dall'algoritmo come ad alto rischio sono trattati dal giudice come se fossero a basso rischio

# Esiste una differenza significativa tra la decisione del giudice e quella del modello di machine learning?

- L'algoritmo prevede che l'**1%** degli imputati più pericolosi avrà un rischio di reiterare il reato del 62,6%, **i giudici rilasciano il 48,5% di questi**
- I dati dimostrano che gli imputati che l'algoritmo aveva predetto essere pericolosi di fatto commettono molti crimini:
  - **Non si presentano in tribunale con un tasso del 56,3%**
  - **Continuano a commettere nuovi crimini con un tasso del 62,7%**
  - **Commettono reati più gravi (omicidio, stupro e rapina) con un tasso del 4,8%**



# Considerazioni...

Vengono sollevate significative sfide econometriche non comuni nelle tipiche applicazioni di apprendimento automatico



Perché i giudici sbagliano? Sono troppo indulgenti?

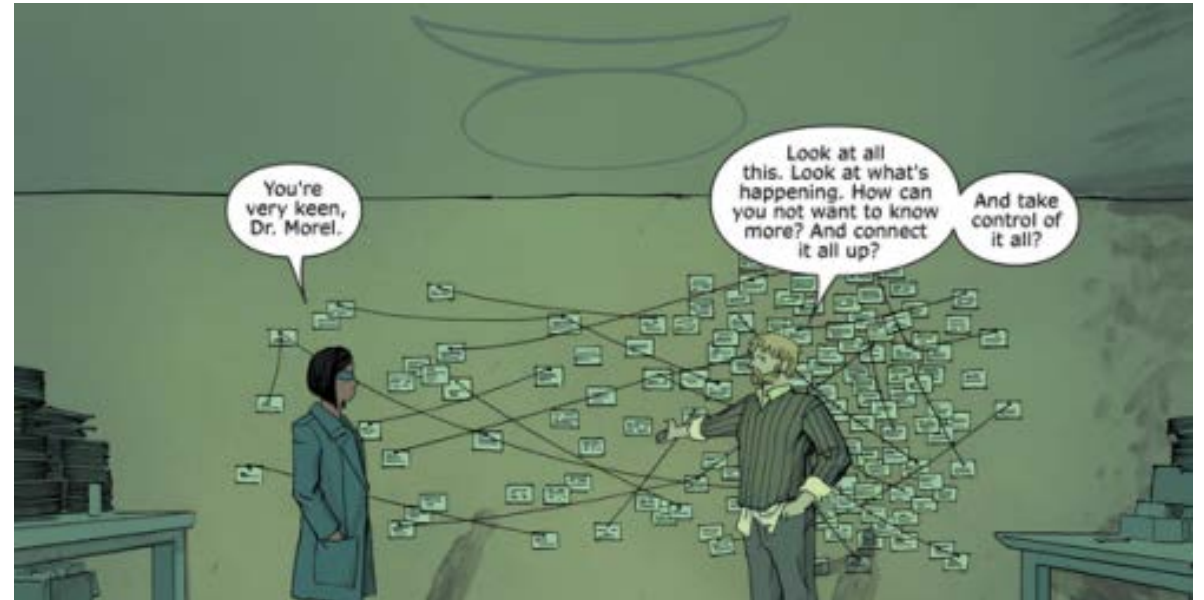


Come sfruttare questa conoscenza per migliorare la decisione?

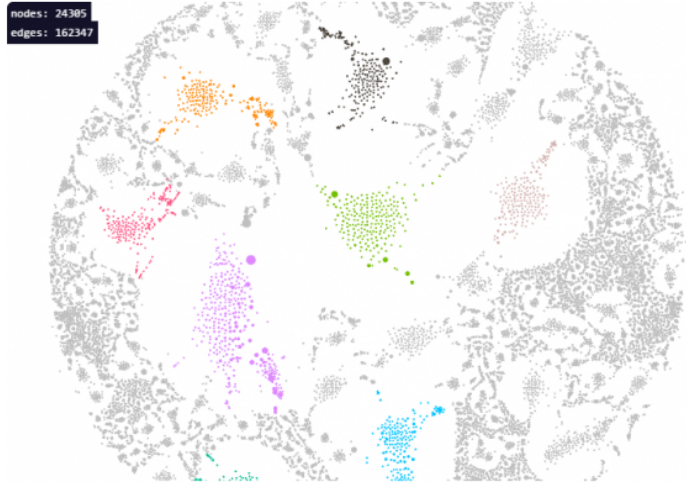
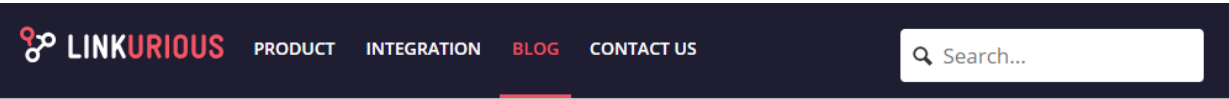


# E la visualizzazione? Osessione per i link

Aspetto particolarmente  
rilevante **anche nel contesto  
giudiziario...**



# Linked Data



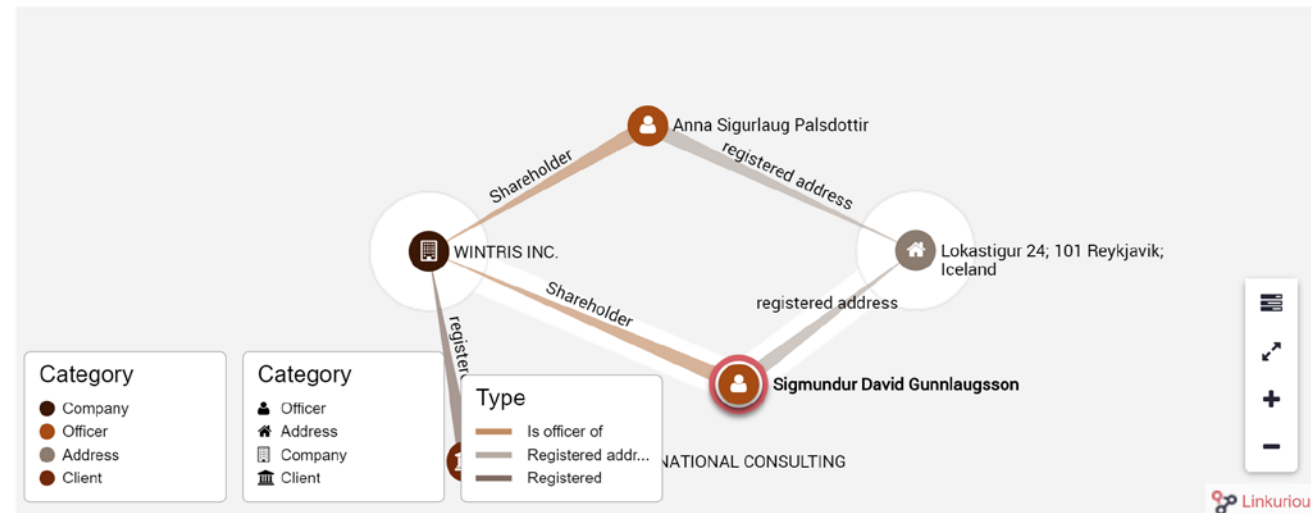
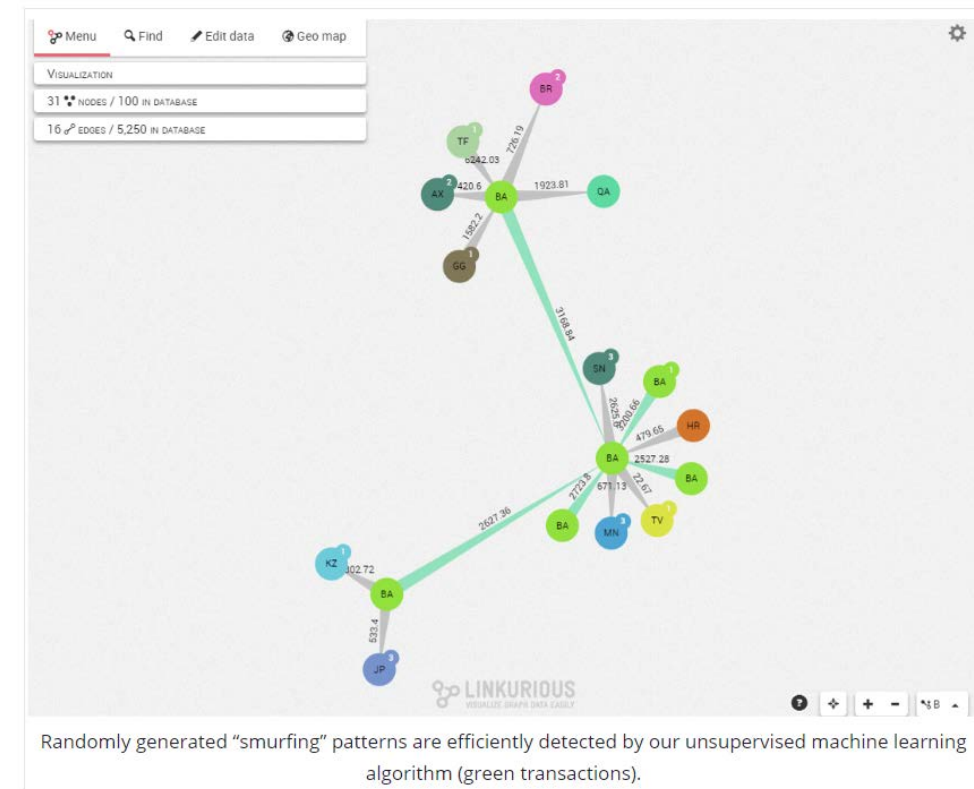
## TECHNOLOGY

### Modular architecture

Ogma makes visualization of large graphs possible by leveraging the full performances of WebGL combined with a modular architecture. The visualization engine is built for WebGL first ; it supports HTML5 Canvas and SVG with the same level of details, though performance are dramatically limited by these rendering technologies.

### Dynamic graphs

The library is designed for graphs with living data where nodes, edges, and their properties can change in real time due to user interactions or updates from data sources.



How Gunnlaugsson hides his secret assets.

<https://linkurio.us/blog/cgi-detect-fraud-prevent-financial-risk/>

# Alcune piattaforme

# RAVEL

BETA

A NEW VIEW ON LEGAL SEARCH

# JUDICATA

Predictice

Au co  
de la

Analysez de  
de justice e

Ravel Law: strumento per la ricerca legale.

Fornisce informazioni dettagliate su come i singoli giudici pensano e governano su questioni particolari.

- Quali opinioni il giudice cita più spesso (e la lingua citata da quelle opinioni)
- Quali giurisdizioni il giudice cita più spesso
- Quali altri giudici il giudice cita più spesso

Gli utenti possono filtrare per parola chiave, tipo di movimento, argomento, statuto o intervallo di date.



Big Law



Law Firms



Government



Corporations

# CrimeMiner



Progetto on-going che coinvolge **esperti provenienti da diverse aree di ricerca**: che vanno dalla scienza sociale all'informatica

# Framework per l'investigazione anche visuale del crimine

---

Estrazione di informazioni da materiale investigativo e gestione visuale

---

Generazione e analisi di network (SNA) attraverso opportune metriche

---

Analisi dell'evoluzione dei gruppi

---

Visualizzazioni georeferenziate dell'organizzazione criminale

---

Machine learning per calcolare la pericolosità

# Punto di partenza

- Studio sociologico che sfrutta le tecniche di SNA per analizzare le caratteristiche di organizzazioni criminali in Sicilia e Campania
- Analisi delle caratteristiche di una rete sociale che emerge dalle intercettazioni telefoniche

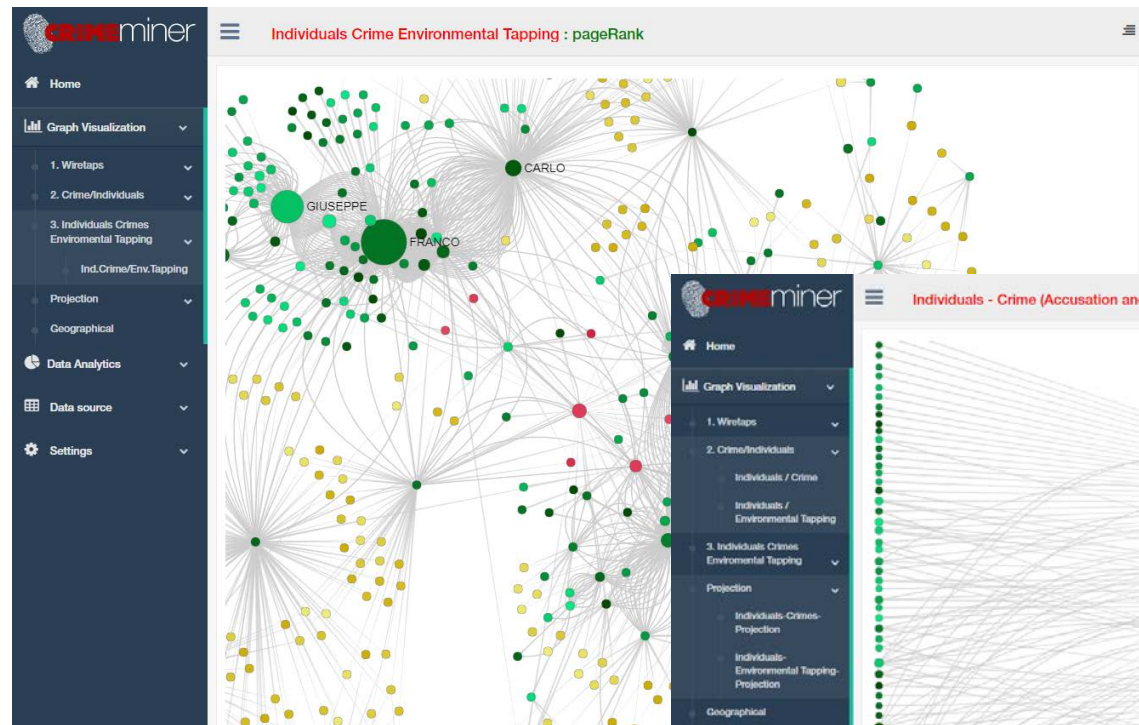


## Caso di studio

Property	Value
Crime class	Organized Crime (Camorra)
Analyzed documents	Remand document
People involved	73
Location	Quindici, Avellino, Italy
Pages of the document	3000
Number of telephone and wiretap	2791
Number of phone under tapping	300

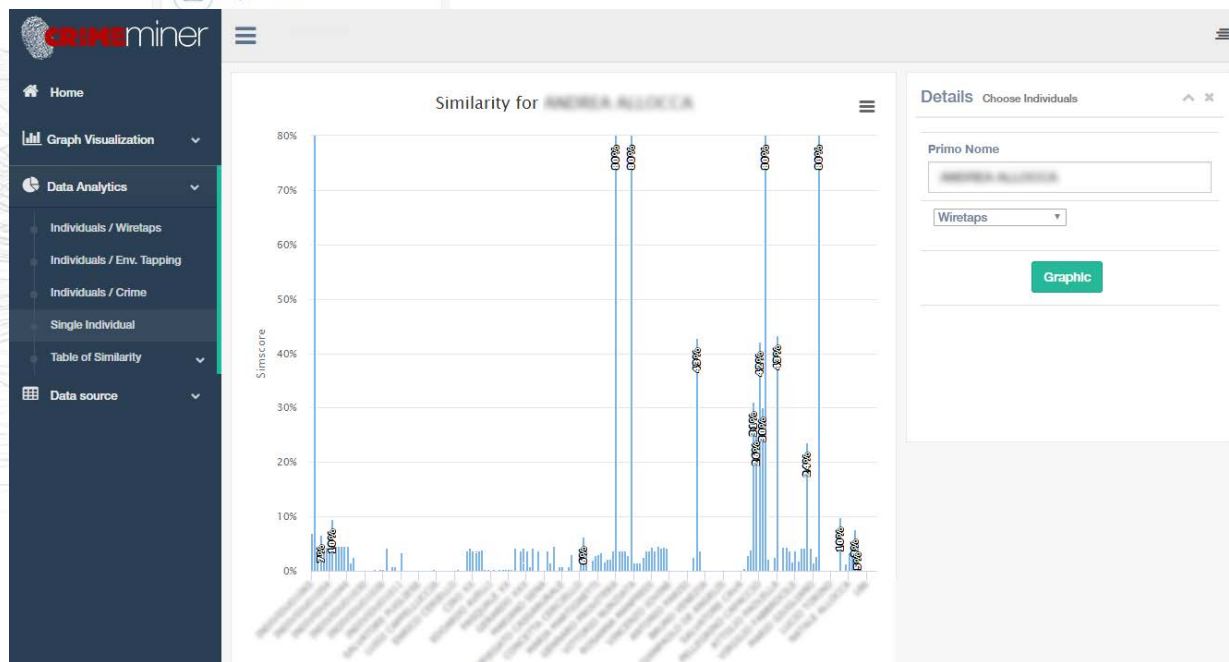
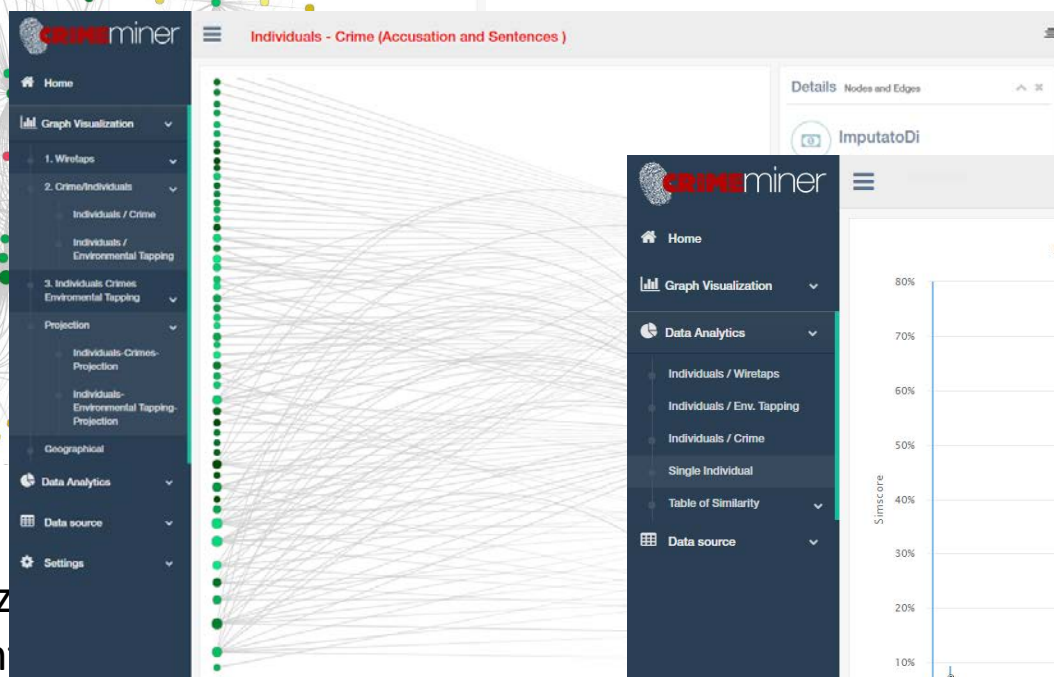


# L'interfaccia



- Community detection
- Reti criminali e identificazione
- Analisi temporale delle interazioni

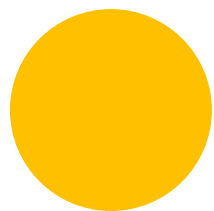
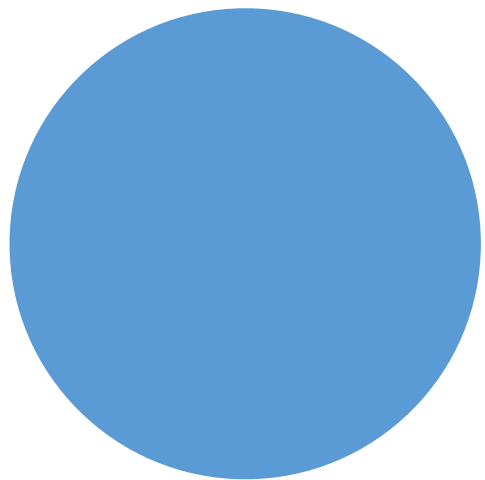
Network bipartite



Similarità tra membri del gruppo

## Take-Home message

- Molti problemi importanti nelle politiche pubbliche sono **problemi predittivi!**
- Gli **algoritmi possono aiutarci** a capire se le decisioni che prendiamo sono in qualche modo viziate (possono aiutarci a fare una diagnosi della ragione del bias)
- Modelli e strumenti informatici maturi anche per il **supporto a decisioni complesse**
- Tante **opportunità inesplorate** ma è necessario un uso sapiente
- I dati sono una **competenza primaria** e non un'attività secondaria che si limita a supportare altri processi



Grazie

