

I Big Data nell'analisi degli ascolti Tv

Marco Catena – RAI Direzione Marketing



Big Data ingombranti? Intelligenze Artificiali non performanti?

Il Supercalcolo: una possibile soluzione

18 Giugno 2020



Auditel



Gli obiettivi

il progresso delle tecniche e delle metodologie di rilevazione per TV, Radio e Digital

analizzare dati per trovare evidenze da utilizzare a sostegno dell'offerta

aud*i*web

La misurazione degli Ascolti TV

Auditel

Auditel è una società che riunisce tutte le componenti del mercato televisivo (editori, investitori pubblicitari, agenzie e centri media) con lo scopo di rilevare in modo oggettivo e imparziale, tutti i giorni, 24 ore su 24, minuto per minuto, la fruizione dell'offerta televisiva e fornire la currency di mercato.



AERANTI-CORALLO

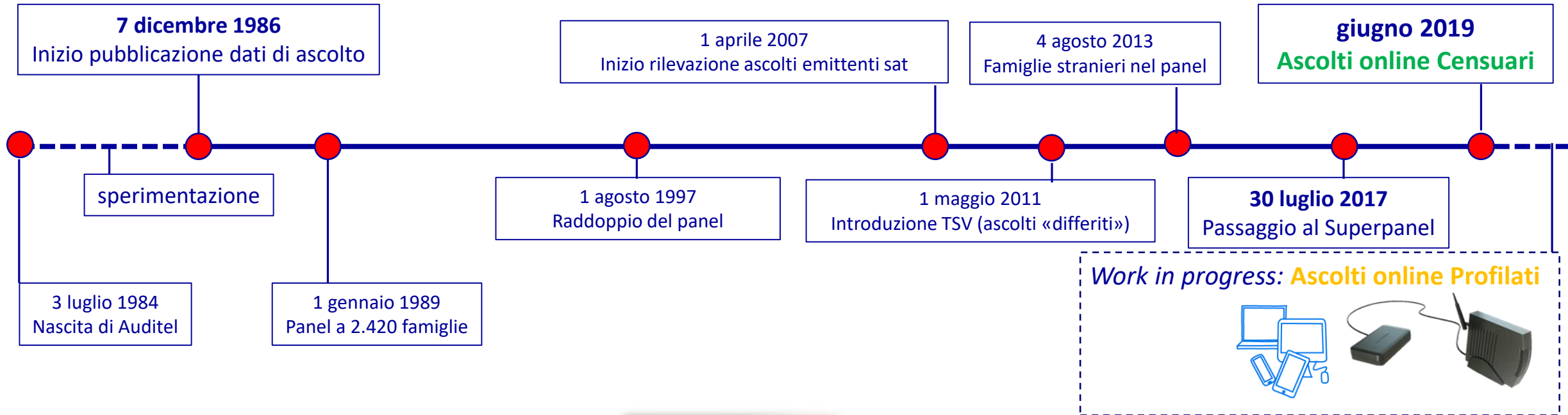
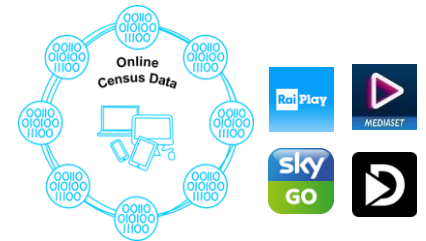
Le imprese radiotelevisive locali, satellitari e via internet



430 Emittenti rilevate - 190+ Editori rilevati

Quantità di dati e Complessità crescenti

Auditel

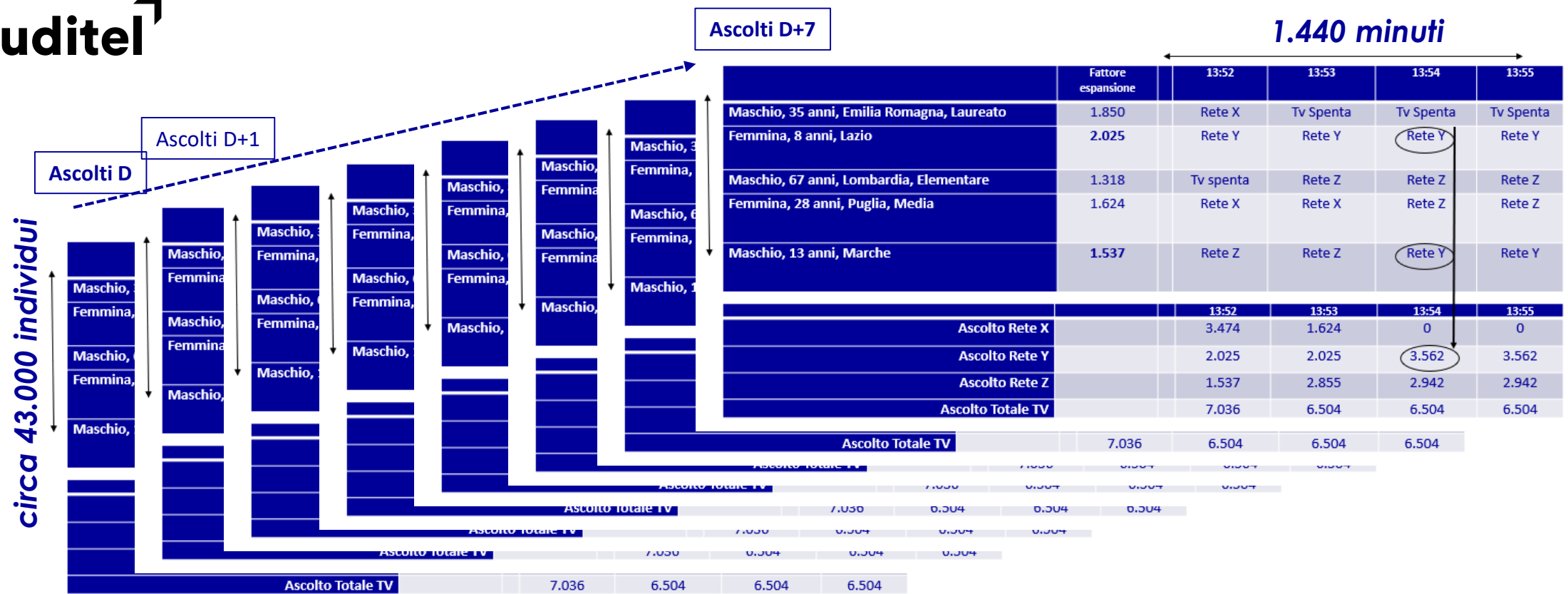


Numerosità Famiglie Panel



La quantità di dati

Auditel



La misurazione della visione differita entro 7 gg (TSV) fa sì che la tabella degli ascolti giornalieri si moltiplichi

(43.000 individui x 11.520 atti di ascolto 1'x1') + (info sociodemo) + (orari programmazione editoriale e pubblicitaria)

≈ 60 GB di dati di ascolto/anno su 10.000 file/anno

La collaborazione

- ❑ CINECA è un soggetto che, operando a supporto dell'ambito accademico e delle imprese, è in grado di mettere a disposizione
 - le potenzialità del Supercalcolo applicate alla ricerca
 - le competenze nell'ambito dell'analisi dei dati e dei modelli statistici, utili per una sintesi delle informazioni adeguata ai fini delle analisi di marketing
- ❑ flessibilità e capacità di far fronte, con rapidità e tempestività, alla produzione di business insights a supporto delle attività editoriali
- ❑ parte attiva nell'individuazione di spunti, progetti e attività di ricerca scientificamente rilevanti nel contesto di misurazione delle audience e dell'analisi dei risultati e, soprattutto, realmente azionabili



Le tipologie di analisi svolte



Clusterizzazione del pubblico

Analisi dei Flussi di ascolto

Stime e previsioni

In sintesi: nuove variabili di caratterizzazione degli ascoltatori per approcci originali alla lettura dei dati

Case study: Indicatore di valutazione della qualità dell'ascolto

La qualità della visione come sintesi dei comportamenti di ascolto

Idee di base:

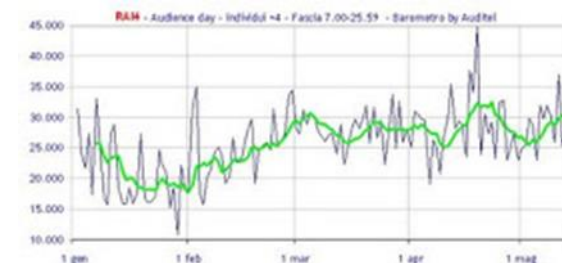
Il successo di un programma televisivo è funzione di:

- **QUANTE persone l'hanno scelto** (= ascolto)
- **COME queste scelte si sono manifestate:**
 - individui con visione continuativa (o con stacchi solo durante i break pubblicitari);
 - individui all'ascolto di altre reti che cambiano canale per seguirlo;
 - individui che sono sul canale anche prima del suo inizio o nei primi minuti di trasmissione;
 - individui che hanno cambiato canale durante un break e che tornano in visione immediatamente prima/dopo il termine della pubblicità

La metodologia:

L'analisi su un periodo consistente dei comportamenti di ascolto dei singoli panelisti Auditel ha permesso di:

- 1) confermare questi fattori come discriminanti nella determinazione della qualità della visione del programma
- 2) creare dei modelli matematici per tradurre i singoli aspetti della visione televisiva in valori numerici crescenti al crescere della caratteristica esaminata, ovvero in fattori
- 3) sintetizzare i diversi fattori in indici, riassunti a loro volta in un unico valore che rappresenta l'indicatore sintetico di "qualità dell'ascolto"



Indicatore di valutazione della qualità dell'ascolto: i 4 indici considerati sulla base di 11 fattori elementari

<p>INTERESSE</p>	<p>misura quanta parte della trasmissione viene vista (rispetto alla sua durata totale); la continuità di visione e la sua capacità di catturare l'attenzione anche degli ascoltatori di altre reti.</p>	<table border="1"> <thead> <tr> <th>fattore</th> <th>cosa esprime</th> </tr> </thead> <tbody> <tr> <td>Share Individuale (ISI)</td> <td>quota del tempo che ogni ascoltatore ha dedicato alla trasmissione rispetto alla visione del mezzo</td> </tr> <tr> <td>Permanenza Individuale (IPI)</td> <td>quota di visione della trasmissione rispetto alla sua durata</td> </tr> <tr> <td>Coinvolgimento (ICO)</td> <td>quota di minuti visti consecutivi rispetto al tempo totale dedicato alla trasmissione</td> </tr> <tr> <td>Cattura (ICA)</td> <td>quota di minuti visti mediamente ad ogni ingresso rispetto al tempo totale dedicato alla trasmissione</td> </tr> </tbody> </table>	fattore	cosa esprime	Share Individuale (ISI)	quota del tempo che ogni ascoltatore ha dedicato alla trasmissione rispetto alla visione del mezzo	Permanenza Individuale (IPI)	quota di visione della trasmissione rispetto alla sua durata	Coinvolgimento (ICO)	quota di minuti visti consecutivi rispetto al tempo totale dedicato alla trasmissione	Cattura (ICA)	quota di minuti visti mediamente ad ogni ingresso rispetto al tempo totale dedicato alla trasmissione
fattore	cosa esprime											
Share Individuale (ISI)	quota del tempo che ogni ascoltatore ha dedicato alla trasmissione rispetto alla visione del mezzo											
Permanenza Individuale (IPI)	quota di visione della trasmissione rispetto alla sua durata											
Coinvolgimento (ICO)	quota di minuti visti consecutivi rispetto al tempo totale dedicato alla trasmissione											
Cattura (ICA)	quota di minuti visti mediamente ad ogni ingresso rispetto al tempo totale dedicato alla trasmissione											
<p>ATTRATTIVA</p>	<p>misura gli individui in visione nella "parte iniziale" della trasmissione (il primo quarto) e il loro peso sull'ascolto dell'intero programma.</p>	<table border="1"> <thead> <tr> <th>fattore</th> <th>cosa esprime</th> </tr> </thead> <tbody> <tr> <td>Quelli che Aspettano (ASP) e (ASC_ASP)</td> <td>Contatti: individui già sintonizzati sulla rete (nei 10 minuti prima dell'inizio della trasmissione) e che vedono in modo consistente la parte iniziale del programma (più della metà) Contributo all'ascolto: loro contributo all'ascolto totale della trasmissione</td> </tr> <tr> <td>Quelli che Entrano (ENT) e (ASC_ENT)</td> <td>Contatti: individui che si sintonizzano sulla rete solo dopo l'inizio della trasmissione ma che ne vedono in modo consistente la parte iniziale (più della metà) Contributo all'ascolto: loro contributo all'ascolto totale della trasmissione</td> </tr> </tbody> </table>	fattore	cosa esprime	Quelli che Aspettano (ASP) e (ASC_ASP)	Contatti: individui già sintonizzati sulla rete (nei 10 minuti prima dell'inizio della trasmissione) e che vedono in modo consistente la parte iniziale del programma (più della metà) Contributo all'ascolto: loro contributo all'ascolto totale della trasmissione	Quelli che Entrano (ENT) e (ASC_ENT)	Contatti: individui che si sintonizzano sulla rete solo dopo l'inizio della trasmissione ma che ne vedono in modo consistente la parte iniziale (più della metà) Contributo all'ascolto: loro contributo all'ascolto totale della trasmissione				
fattore	cosa esprime											
Quelli che Aspettano (ASP) e (ASC_ASP)	Contatti: individui già sintonizzati sulla rete (nei 10 minuti prima dell'inizio della trasmissione) e che vedono in modo consistente la parte iniziale del programma (più della metà) Contributo all'ascolto: loro contributo all'ascolto totale della trasmissione											
Quelli che Entrano (ENT) e (ASC_ENT)	Contatti: individui che si sintonizzano sulla rete solo dopo l'inizio della trasmissione ma che ne vedono in modo consistente la parte iniziale (più della metà) Contributo all'ascolto: loro contributo all'ascolto totale della trasmissione											
<p>RESISTENZA AI CAMBI CANALE</p>	<p>misura la capacità di trattenere gli ascoltatori in relazione anche alla situazione sulle altre reti</p>	<table border="1"> <thead> <tr> <th>fattore</th> <th>cosa esprime</th> </tr> </thead> <tbody> <tr> <td>Cambi canale → Altre reti (ICCU1)</td> <td>rapporto tra cambi canale effettuati e potenziali</td> </tr> <tr> <td>Cambi canale → Tv spento (ICCU2)</td> <td>rapporto tra atti di spegnimento effettuati e potenziali</td> </tr> <tr> <td>Stabilità della rete (ICCU3)</td> <td>raffronto tra: Cambi canale effettuati e potenziali sulle altre reti Cambi canali effettuati e potenziali sulla rete in esame</td> </tr> </tbody> </table>	fattore	cosa esprime	Cambi canale → Altre reti (ICCU1)	rapporto tra cambi canale effettuati e potenziali	Cambi canale → Tv spento (ICCU2)	rapporto tra atti di spegnimento effettuati e potenziali	Stabilità della rete (ICCU3)	raffronto tra: Cambi canale effettuati e potenziali sulle altre reti Cambi canali effettuati e potenziali sulla rete in esame		
fattore	cosa esprime											
Cambi canale → Altre reti (ICCU1)	rapporto tra cambi canale effettuati e potenziali											
Cambi canale → Tv spento (ICCU2)	rapporto tra atti di spegnimento effettuati e potenziali											
Stabilità della rete (ICCU3)	raffronto tra: Cambi canale effettuati e potenziali sulle altre reti Cambi canali effettuati e potenziali sulla rete in esame											
<p>RESISTENZA ALLA PUBBLICITA'</p>	<p>misura la permanenza all'ascolto della rete nell'intorno degli intervalli pubblicitari</p>	<table border="1"> <thead> <tr> <th>fattore</th> <th>cosa esprime</th> </tr> </thead> <tbody> <tr> <td>Richiamo (IRG)</td> <td>Contributo all'ascolto di chi, in visione prima del break (tutti e 5 i minuti prima dello stesso) e' in visione sulla rete anche prima della fine del break (nei 3 minuti che precedono la sua fine) - dato medio su tutti i break</td> </tr> <tr> <td>Mantenimento dopo la pubblicità (IMDP)</td> <td>Contributo all'ascolto di chi, in visione prima del break (tutti e 5 i minuti prima dello stesso) e' in visione sulla rete anche subito dopo la fine del break (in almeno 1 dei 5 minuti successivi alla ripresa della trasmissione) - dato medio su tutti i break</td> </tr> <tr> <td>Rientro dopo la pubblicità (IRDPA)</td> <td>Contributo all'ascolto di chi, in visione prima del break (in almeno 1 dei 3 minuti precedenti) e' in visione sulla rete anche subito dopo la fine del break (in almeno 1 dei 5 minuti successivi alla ripresa della trasmissione) - dato medio su tutti i break</td> </tr> </tbody> </table>	fattore	cosa esprime	Richiamo (IRG)	Contributo all'ascolto di chi, in visione prima del break (tutti e 5 i minuti prima dello stesso) e' in visione sulla rete anche prima della fine del break (nei 3 minuti che precedono la sua fine) - dato medio su tutti i break	Mantenimento dopo la pubblicità (IMDP)	Contributo all'ascolto di chi, in visione prima del break (tutti e 5 i minuti prima dello stesso) e' in visione sulla rete anche subito dopo la fine del break (in almeno 1 dei 5 minuti successivi alla ripresa della trasmissione) - dato medio su tutti i break	Rientro dopo la pubblicità (IRDPA)	Contributo all'ascolto di chi, in visione prima del break (in almeno 1 dei 3 minuti precedenti) e' in visione sulla rete anche subito dopo la fine del break (in almeno 1 dei 5 minuti successivi alla ripresa della trasmissione) - dato medio su tutti i break		
fattore	cosa esprime											
Richiamo (IRG)	Contributo all'ascolto di chi, in visione prima del break (tutti e 5 i minuti prima dello stesso) e' in visione sulla rete anche prima della fine del break (nei 3 minuti che precedono la sua fine) - dato medio su tutti i break											
Mantenimento dopo la pubblicità (IMDP)	Contributo all'ascolto di chi, in visione prima del break (tutti e 5 i minuti prima dello stesso) e' in visione sulla rete anche subito dopo la fine del break (in almeno 1 dei 5 minuti successivi alla ripresa della trasmissione) - dato medio su tutti i break											
Rientro dopo la pubblicità (IRDPA)	Contributo all'ascolto di chi, in visione prima del break (in almeno 1 dei 3 minuti precedenti) e' in visione sulla rete anche subito dopo la fine del break (in almeno 1 dei 5 minuti successivi alla ripresa della trasmissione) - dato medio su tutti i break											

Indicatore di valutazione della qualità dell'ascolto: i pesi degli indici

L'indicatore di valutazione della qualità dell'ascolto è dato dalla media ponderata dei quattro indici, dove i pesi:

- ✓ sono determinati in base alla "quantità di informazione" spiegata (pesi più alti per gli indici più significativi)
- ✓ non sono univoci: poiché la fruizione tv varia in funzione dei **generi** e delle **fasce di trasmissione**, il contributo "informativo" di ciascun fattore ha un valore diverso in base a questi elementi

Qualità Ascolto = $p_1 \times \text{INTERESSE} + p_2 \times \text{ATTRATTIVA} + p_3 \times \text{RESISTENZA CAMBI CANALE} + p_4 \times \text{RESISTENZA PUBBLICITA'}$

ESEMPIO:

INTRATTENIMENTO

Fascia Mattina

$0,33 \times \text{Interesse} + 0,23 \times \text{Attrattiva} + 0,22 \times \text{Resistenza Cambi Canale} + 0,22 \times \text{Resistenza alla Pubblicità}$

Fascia Prima Serata

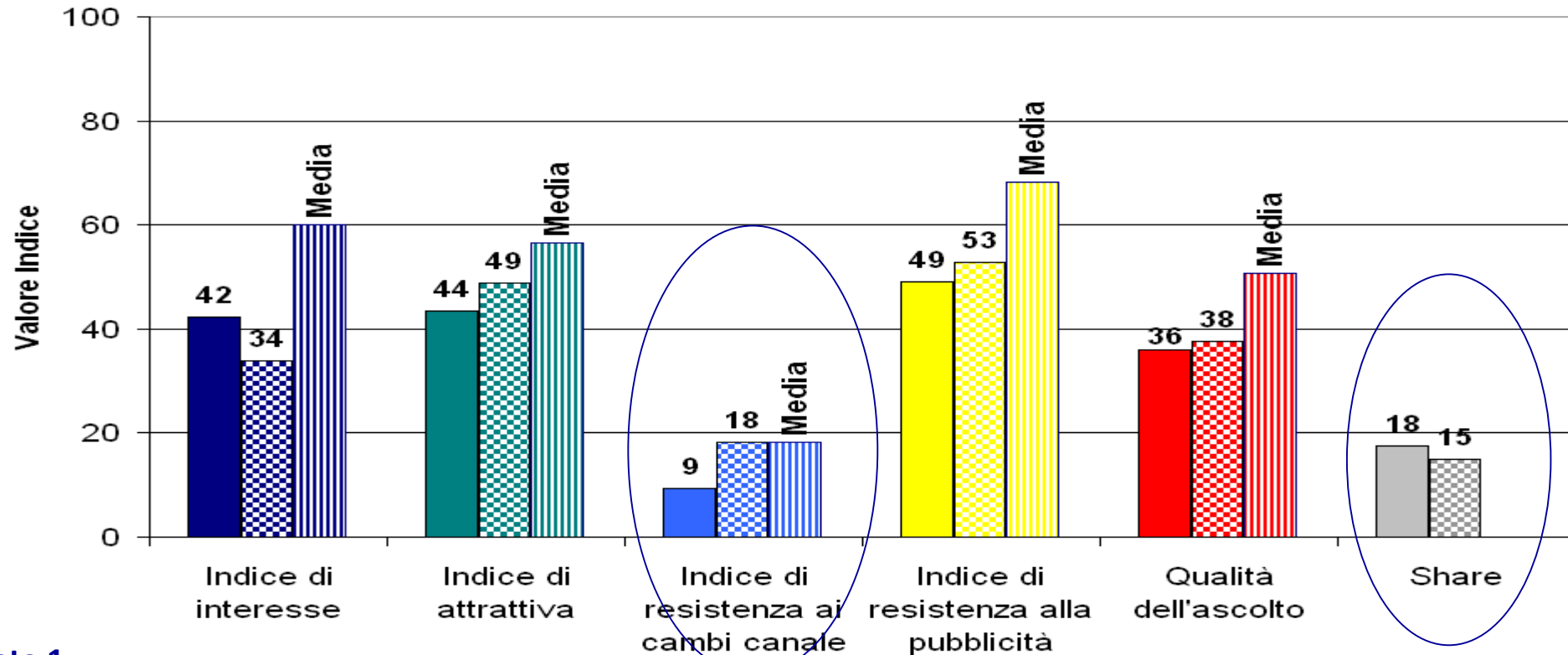
$0,30 \times \text{Interesse} + 0,22 \times \text{Attrattiva} + 0,19 \times \text{Resistenza Cambi Canale} + 0,29 \times \text{Resistenza alla Pubblicità}$

L'indicatore ha un valore predittivo

Dall'analisi degli indici di una trasmissione appena cominciata posso dedurre il successo

Programma P
Genere Intrattenimento

Indici su Base 0-100



-  Puntata 1
-  Puntata 2
-  Media intrattenimento – Rete X – Prima Serata

Case study: Individuazione del Forte Ascoltatore

uso delle metriche «tradizionali»



Forte Ascoltatore di Rete_X:

preferisce trascorrere il suo tempo di visione su Rete_X piuttosto che sulle altre reti.

Indicatore1 = rapporto degli share

Share *individuale* sulla RETE_X
>
Share medio della RETE_X

Indicatore2 = rapporto dei minuti visti

Minuti visti *individuali* sulla
RETE_X
>
Minuti medi della RETE_X

Il comportamento di un individuo è valutato indipendentemente da quello degli altri

Limite di questo approccio: Multi-assegnazione

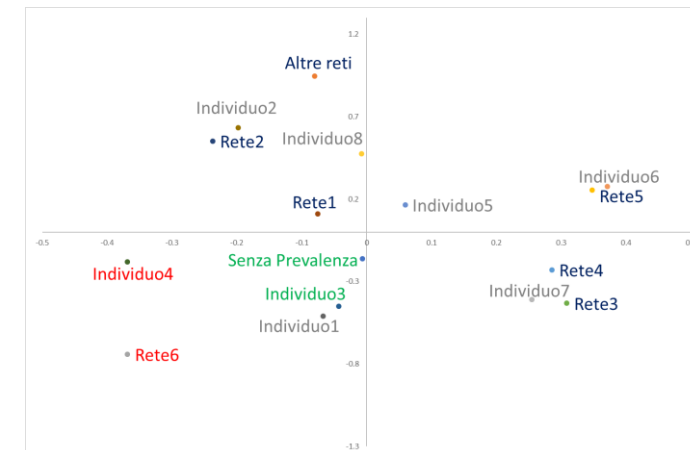
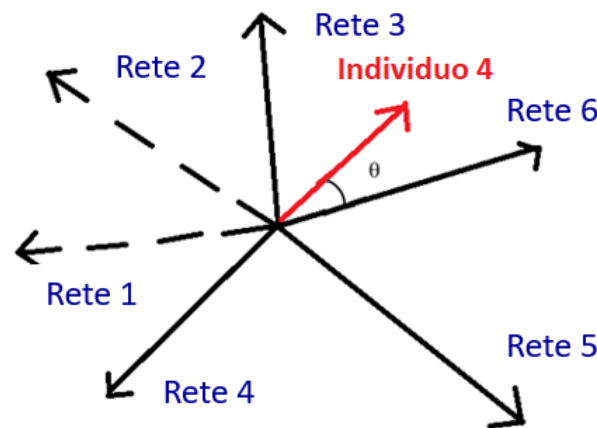
Individuazione del Forte Ascoltatore

uso di tecniche di analisi statistica multivariata

L'impiego di tecniche di tipo geometrico permette di tenere conto *simultaneamente* di tutte le dinamiche di fruizione degli ascoltatori, considerando il comportamento di visione individuale nella sua totalità confrontandolo con quello medio degli individui su ciascuna rete e sul complesso delle reti contemporaneamente.

Individuo	Rete1	Rete2	Rete3	Rete4	Rete5	Rete6	Altre Reti	Non Ascolto	Cosine of the angle between the row and the assignment column	Nearest variable
1	5	1	2	3	5	8	0	0	0.95	Rete6
2	5	6	0	0	3	0	3	7	0.84	Altre Reti
3	0	0	0	1	0	1	0	22	0.98	Senza prevalenza
4	7	4	2	3	1	6	1	0	0.75	Rete6
5	9	4	2	5	0	0	1	3	0.82	Rete1
6	4	3	1	4	7	2	2	1	0.92	Rete5
7	5	1	5	3	1	2	0	7	0.88	Rete3
8	4	3	1	0	3	0	4	9	0.83	Altre Reti
Totale	39	22	13	19	20	19	11	49		

Proiettando l'individuo in uno spazio N-dimensionale (dove N sono le reti) la rete di cui è «Forte ascoltatore» è quella per la quale il *coseno dell'angolo tra rete è individuo è maggiore*

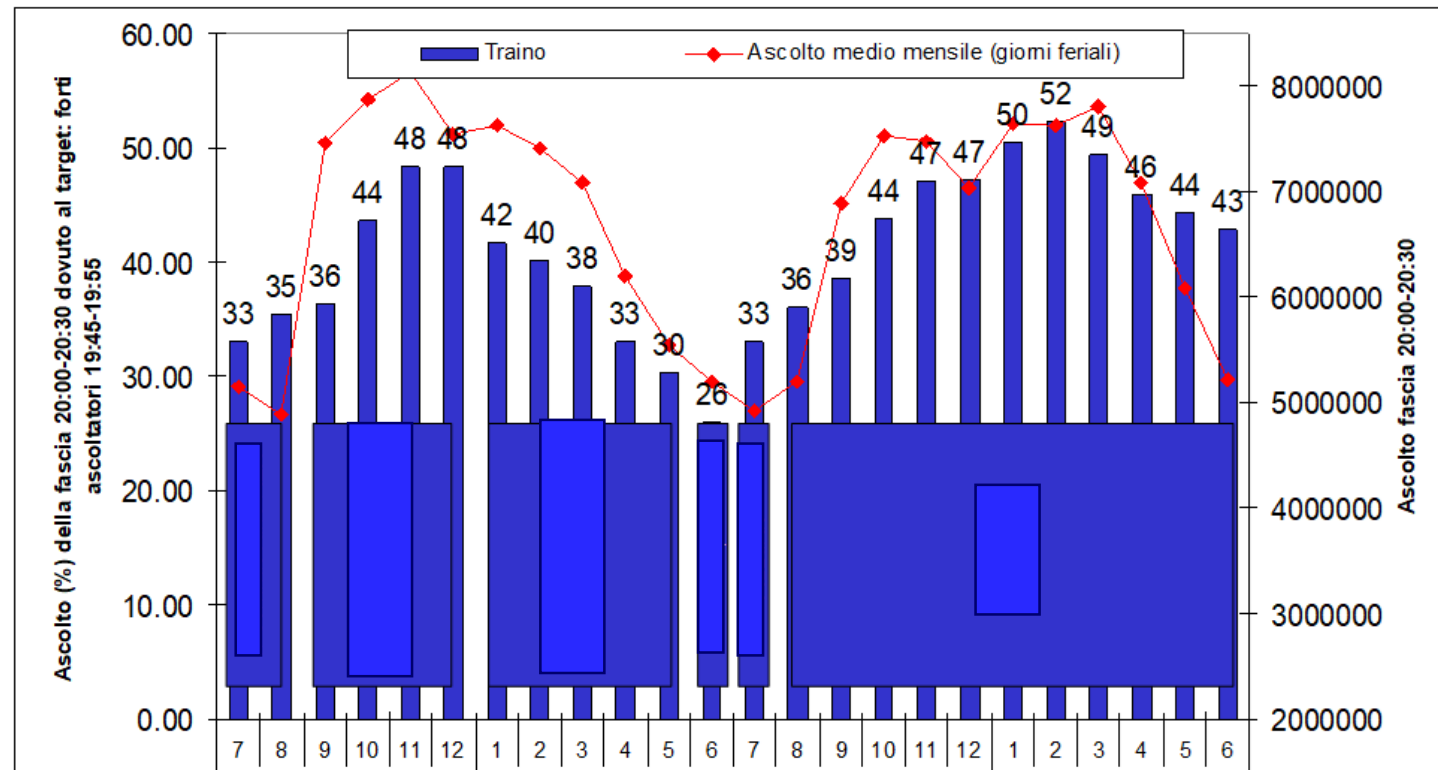


Individuazione del Forte Ascoltatore

un esempio di utilizzo

Analisi del «traino» da una fascia oraria ad un'altra
(in termini di numerosità di ascoltatori e contributo all'ascolto)

L'andamento mensile del traino al TG1 (Giorni feriali, con riferimento alle trasmissioni prevalenti nel mese)

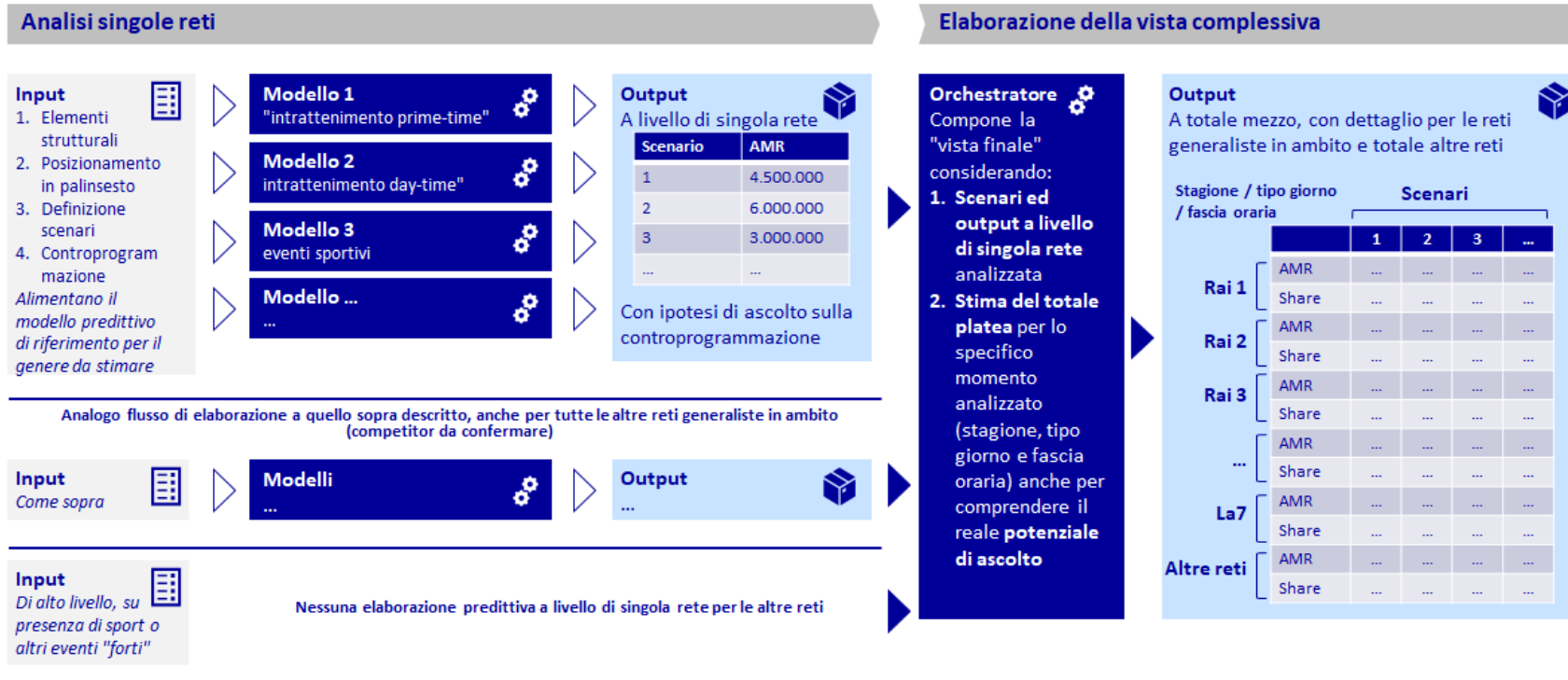


Quali prospettive per il futuro?

Sviluppo di Proof of concept con il principale obiettivo di estrarre ulteriore valore dai dati sia attraverso l'integrazione dei dati Auditel con nuove fonti (interne ed esterne) sia attraverso l'abilitazione di nuovi paradigmi di analisi e di interpretazione.

Proprio in questa direzione, abbiamo intenzione di fare leva sui servizi e sulle competenze che offre CINECA in ambito Big Data per sfruttare appieno le potenzialità del machine learning e dell'intelligenza artificiale.

Proof of concept – Modello stima ascolti attesi



Tale stima potrà essere di supporto:

- nella fase iniziale, di definizione del programma, per la selezione del "mix di ingredienti" (elementi strutturali e di posizionamento) che possa massimizzare il ritorno atteso in termini di ascolto;
- nella fase di definizione operativa del palinsesto, considerando le ipotesi sulle scelte di programmazione/controprogrammazione che potrebbero essere fatte dai principali competitor.



I Big Data nell'analisi degli ascolti Tv

Big Data ingombranti? Intelligenze Artificiali non performanti?

Il Supercalcolo: una possibile soluzione

18 Giugno 2020