



UNIVERSITÀ
DI PISA

Forecast emerging artists success on Last.fm service: a data-driven study

Relatrici:

Prof.essa Anna Monreale

Ph.D. Laura Pollacci

Candidata:

Alexandra Lavinia Bradan

8 maggio 2020

Contenuto

1

Introduzione

Obiettivo della tesi
Rete sociale Last.fm

2

Modello diffusivo e modello predittivo

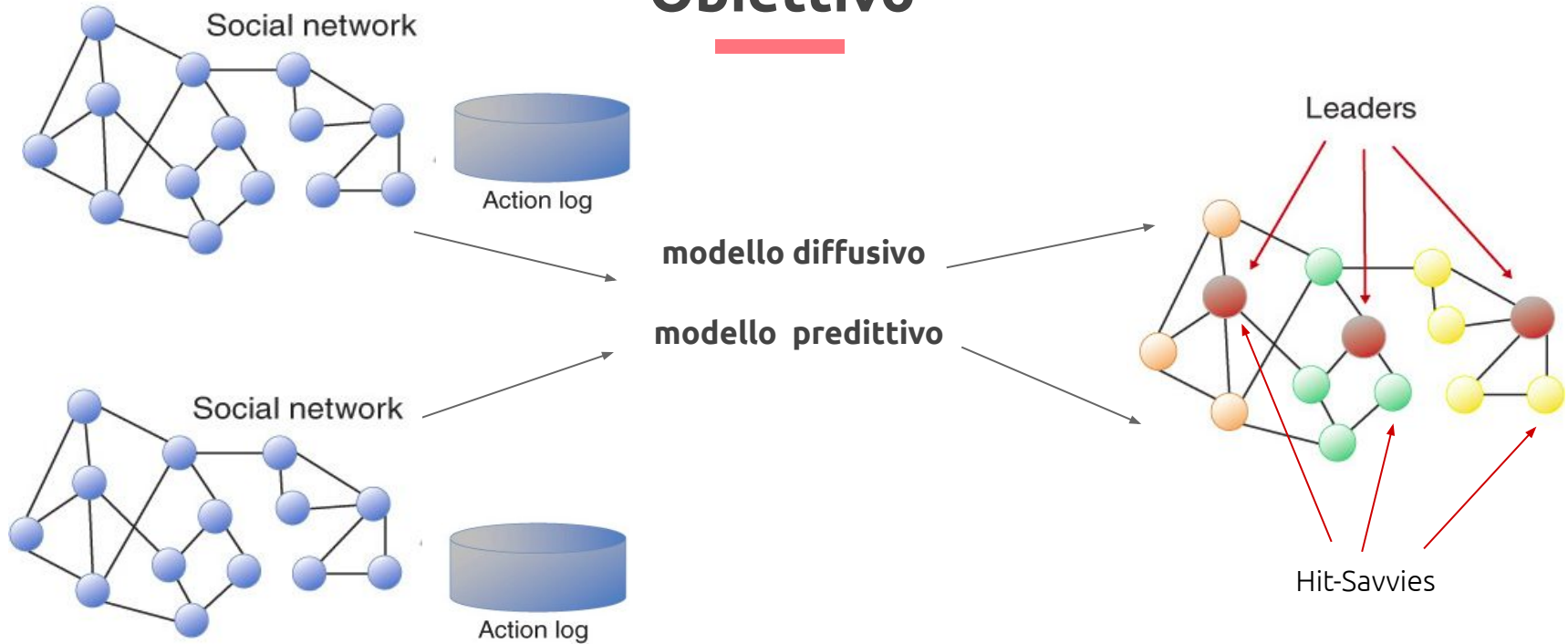
Direct mining distribution model
Threshold predictive model

3

Conclusioni

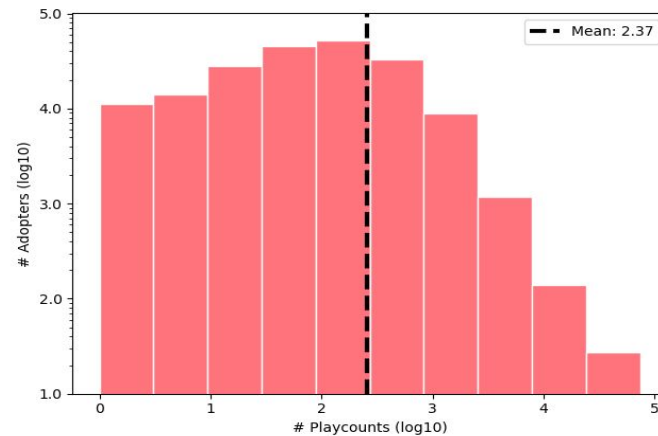
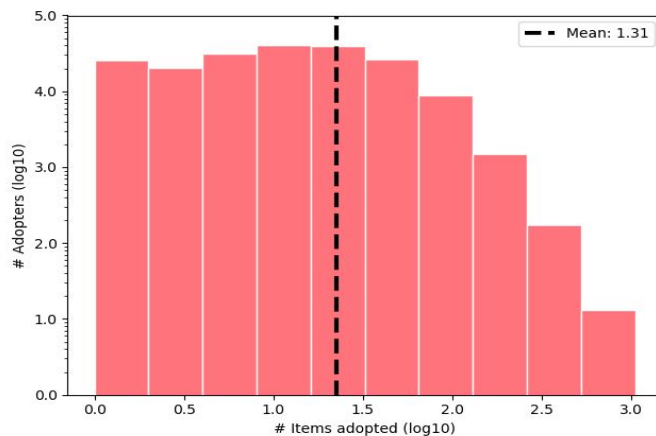
Comparazione tra i due modelli
Osservazioni finali

Obiettivo

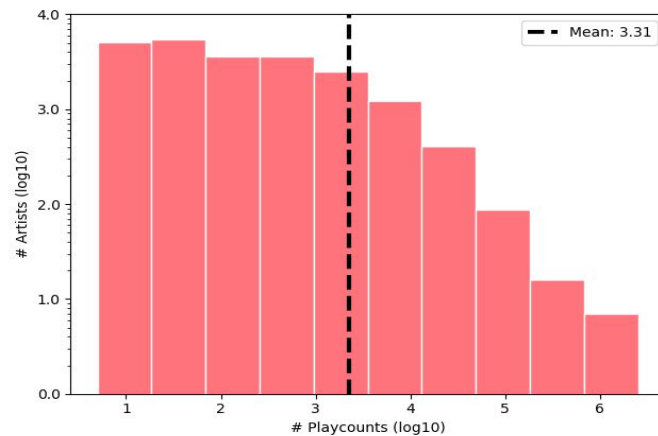
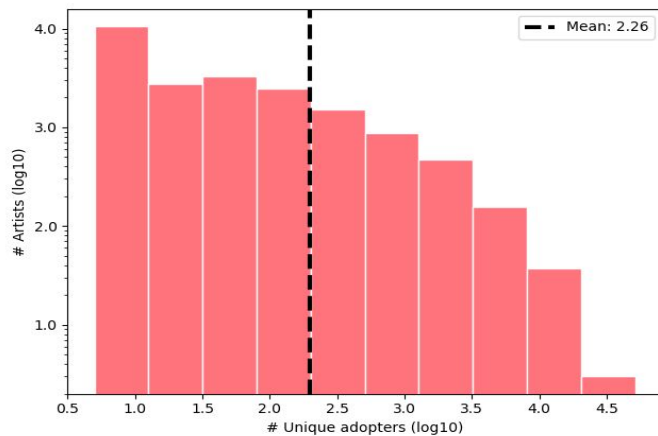
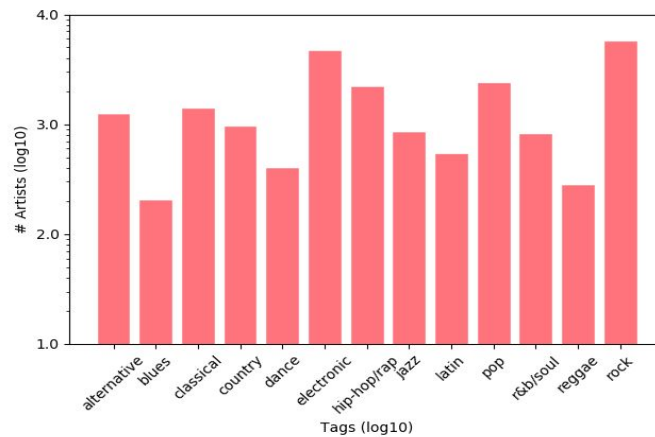




artisti	utenti	azioni	timespan	granularità
22 001	193 409	11 620 917	2 anni*	1 settimana



last.fm



Notazione

G = (V, E) grafo

V = insieme degli utenti/adottatori

E = $\{(u, v) \mid u, v \in V \wedge u \text{ è follower di } v\}$

Ψ = insieme degli oggetti/artisti adottati

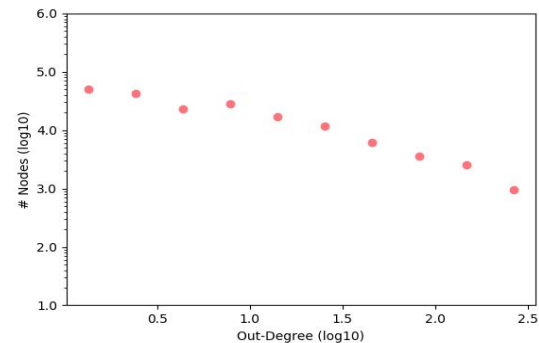
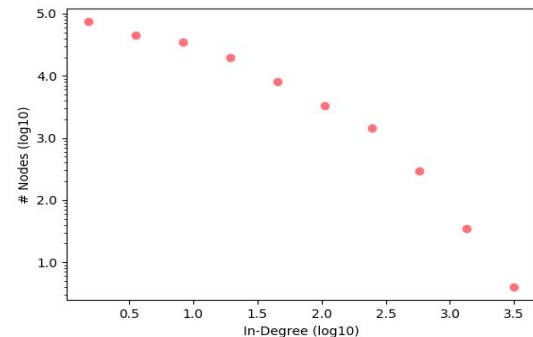
A = insieme delle azioni (granularità settimanale)

S = insieme delle settimane

$t_{i(v, \psi)} = \min_{i=1, \dots, n} t.c. v \in V, \psi \in \Psi, i \in S \wedge \exists a_{v, \psi} \in A$

$\omega(v, \Psi) = \{\psi \mid \psi \in \Psi, v \in V \wedge \exists a_{v, \psi} \in A\}$

$\theta(\psi, V) = \{v \mid v \in V, \psi \in \Psi \wedge \exists a_{v, \psi} \in A\}$



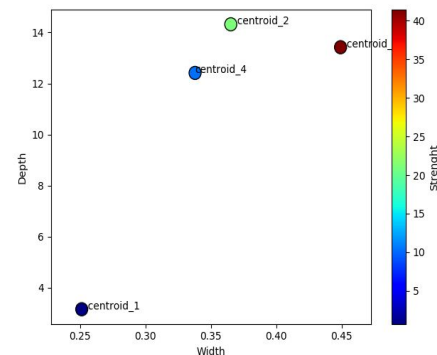
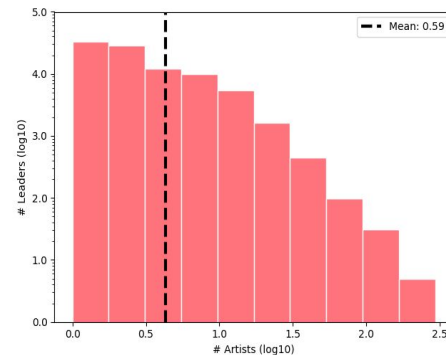
Modello diffusivo



```
leaders = []
diffusion_trees = {}
 $\forall \psi \in \Psi$ :
    G' $\psi$  = new DiGraph()
    retrieve  $\psi$ 's G $\psi$  = (V $\psi$ , E $\psi$ ) s.t. V $\psi$   $\subseteq$  V, E $\psi$   $\subseteq$  E
    for u, v in E $\psi$  :
        if t_v $\psi$   $\leq$  t_u $\psi$  and t_u $\psi$  - t_v $\psi$   $\leq$   $\delta$ :
            G' $\psi$ .add_edge(v, u, weight=t_v $\psi$ )

    remove isolated nodes from G' $\psi$ 
    retrieve each connected component DAG from G' $\psi$ 
    for DAG in G' $\psi$  :
        v = list(topological_sort(DAG))[0]
        leaders.append(v),
        diffusion_trees[v] = {"item":  $\psi$ }
        diffusion_trees[v]["diffusion_tree"] = min_spanning_arborescence(DAG)

return leaders, diffusion_trees
```



Modello predittivo: ranking del successo

Definizione 1

Artisti con più ascoltatori tra i miei seed users

Definizione 2

Artisti con più ascolti tra i miei seed users

Definizione 3

Artisti con più azioni tra i miei seed users

Definizione 4

Artisti con più ascoltatori tra gli utenti di Last.fm

Definizione 5

Artisti con più ascolti tra gli utenti di Last.fm

Definizione 6

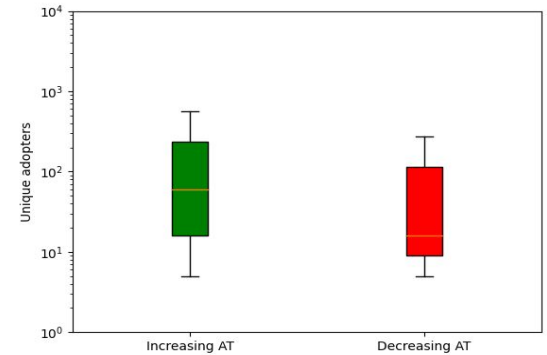
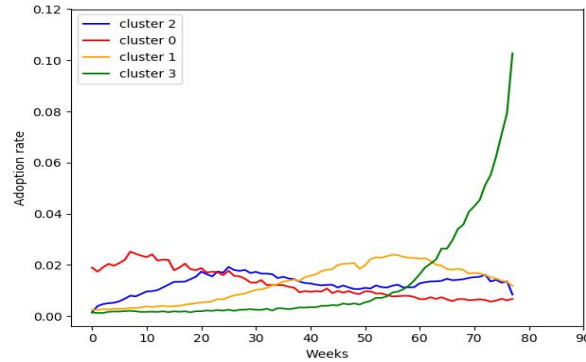
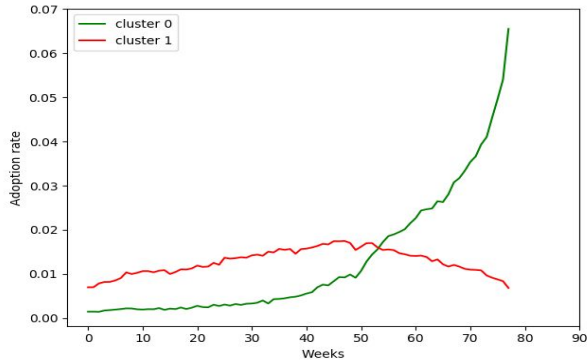
Artisti più ricercati su Google

Definizione 7

Artisti con maggior numero di nuovi ascoltatori crescente nel tempo.

Adoption Trend: dato un oggetto ψ , il suo adoption trend $T(\psi)$ è definito come:

$$T(\psi) = \{ \tau(t_1), \dots, \tau(t_n) \} \text{ dove } \tau(t_i) = |\{v \mid v \in V, \psi \in \omega(v, \Psi) \wedge t_{i(v, \psi)} = t_i\}| / |\theta(\psi, V)| \quad \forall i = 1, \dots, n$$



$$\mathbf{H} = \{\psi \mid \psi \in \Psi \wedge \psi \in \text{Hit}\}$$

$$\mathbf{F} = \{\psi \mid \psi \in \Psi \wedge \psi \in \text{Flop}\}$$

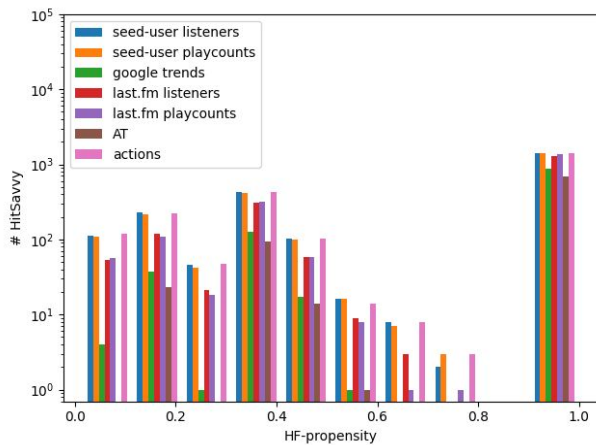
Modello predittivo: indicatori

HF-propensity: dato $v \in V$, $h = |\omega(v, \Psi')|$ con $\Psi' \subseteq H$, $k = |\omega(v, \Psi'')|$ con $\Psi'' \subseteq H \setminus \Psi'$ (late Hits), $f = |\omega(v, \Psi''')|$ con $\Psi''' \subseteq F$, la sua HF-propensity è definita come:

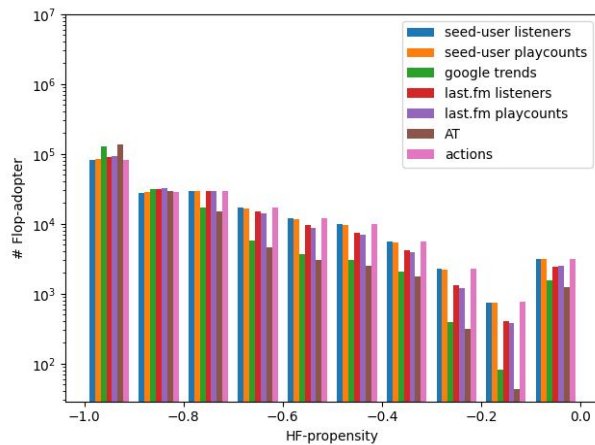
$$\text{HF}(v) = h - k - f / h + k + f$$

$$\text{HF}(v) \in [-1, 1]$$

Hit-Savvies = $\{v \mid v \in V \wedge \text{HF}(v) > 0\}$



Flop-adopters = $\{v \mid v \in V \wedge \text{HF}(v) < 0\}$



Weighted Multi-Set Coverage (WMSC)

$\mathbf{X} = \text{sorted}(\text{Hit-Savvies}^*)$

$\mathbf{Y} = \bigcup_{i=1, \dots, |\mathbf{X}|} \omega(x_i, H^{**})$

$\mathbf{E} = \{(x, y) \mid x \in \mathbf{X}, y \in \mathbf{Y}\}$

$\mathbf{BG} = (\mathbf{X}, \mathbf{Y}, \mathbf{E})$ grafo bipartito

*Flop-adopters

**F

$$\min x_1, \dots, x_{|\mathbf{X}|} \quad \sum x_i$$

$$\text{subject to} \quad \forall i, j: x_i = e_{ij}$$

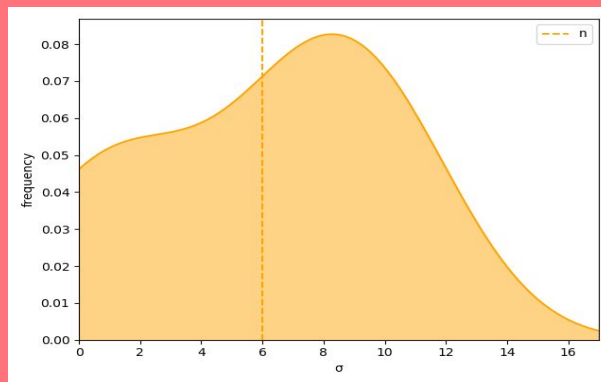
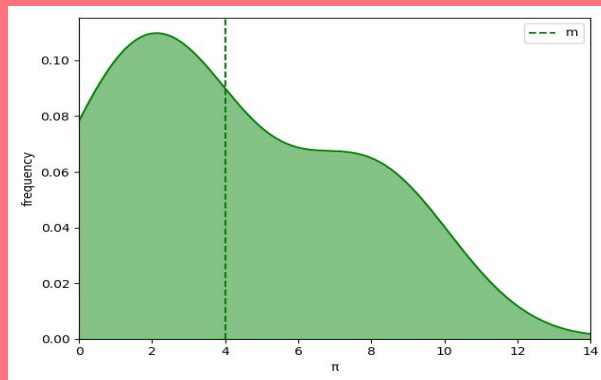
$$\forall j: \sum e_{ij} \geq y_j \beta$$

$$\sum y_j \geq \alpha |\mathbf{Y}|$$

where $x_i, y_j, e_{ij} \in \{0, 1\}, \alpha, \beta \in (0, 1]$

Hitters = $X' \subseteq \text{sorted}(\text{Hit-Savvies})$

Floppers = $X' \subseteq \text{sorted}(\text{Flop-adopters})$



$$\pi(\psi) = |\theta(\psi, \text{Hitters})| \text{ con } \psi \in H$$

$$\sigma(\psi) = |\theta(\psi, \text{Floppers})| \text{ con } \psi \in F$$

Modello predittivo con WMSC

ψ_* = nuovo artista

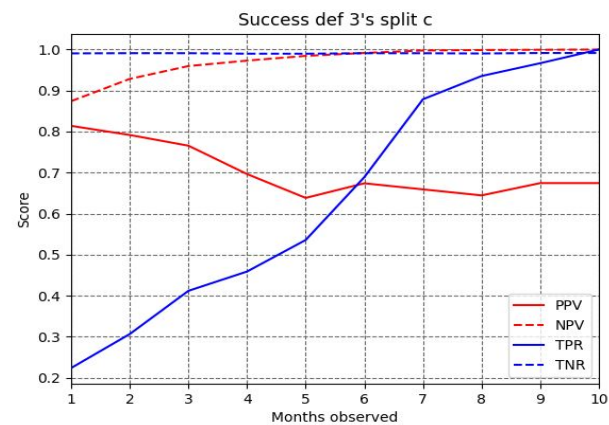
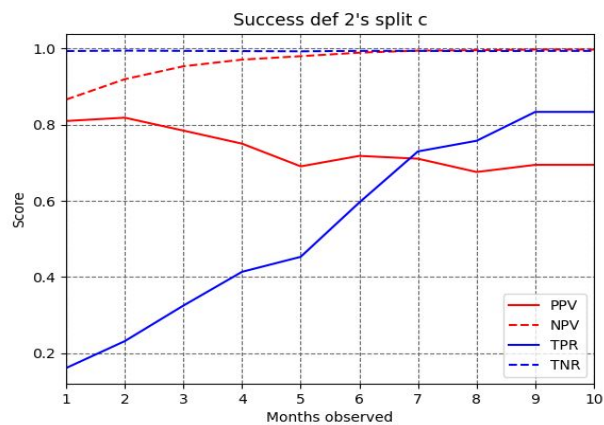
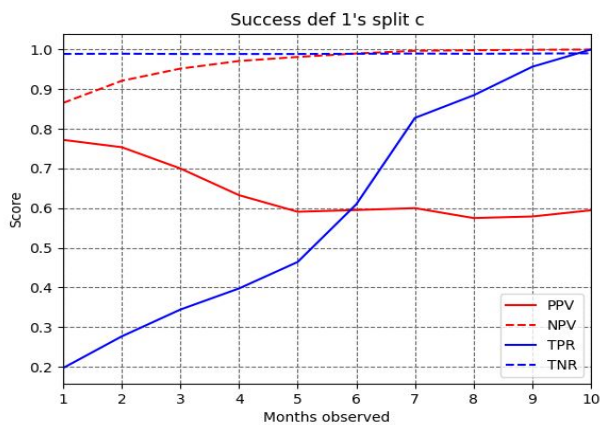
$\theta(\psi_*, V)$ = adottatori di ψ_* in un range limitato di tempo dalla sua prima comparsa

Hits-classifier H_p : ritorna una classe positiva (Hit) sse $n(\psi_*) > m$, con $n(\psi_*)$ calcolato su $\theta(\psi_*, V \cap \text{Hitters})$.

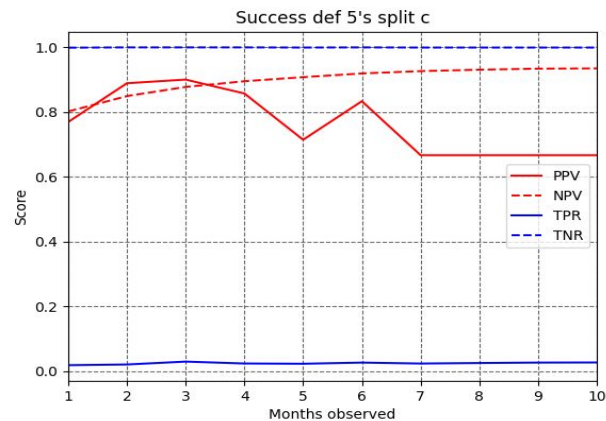
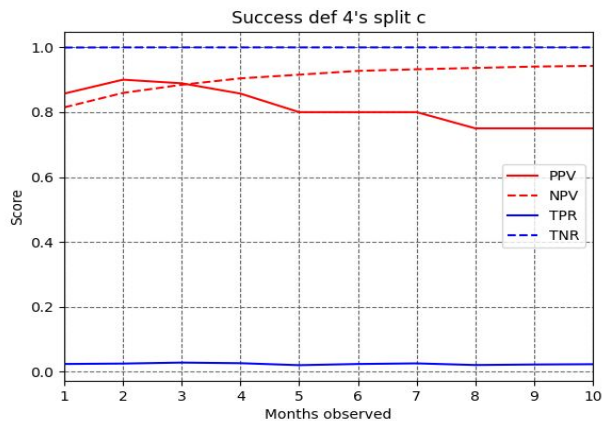
Flops-classifier F_p : ritorna una classe positiva (Flop) sse $\sigma(\psi_*) > n$, con $\sigma(\psi_*)$ calcolato su $\theta(\psi_*, V \cap \text{Flopper})$.

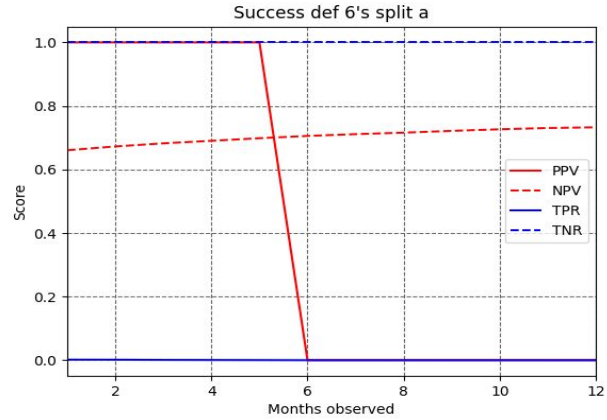
Hits&Flops meta-classifier: dati H_p and F_p , il rule based meta-classifier è così definito:

1. se $H_p(\psi_*) \rightarrow \text{Hit}$ e $F_p(\psi_*) \rightarrow \text{Hit}$, ψ_* è un futuro Hit;
2. se $H_p(\psi_*) \rightarrow \text{Flop}$ e $F_p(\psi_*) \rightarrow \text{Flop}$, ψ_* è un futuro Flop;
3. se $H_p(\psi_*) \rightarrow \text{Hit}$ e $F_p(\psi_*) \rightarrow \text{Flop}$:
 - (a) se $(n(\psi_*) - m) > (\sigma(\psi_*) - n) \rightarrow \psi_*$ è un futuro Hit;
 - (b) se $(n(\psi_*) - m) = (\sigma(\psi_*) - n)$ le adozioni osservate non sono sufficiente per fare una previsione;
 - (c) altrimenti, ψ_* è un futuro Flop;
4. altrimenti, le adozioni osservate non sono sufficiente per fare una previsione.



last.fm



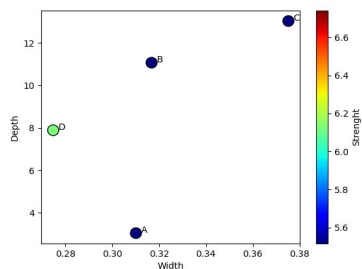


Comparazione Hit-Savvies & Leaders

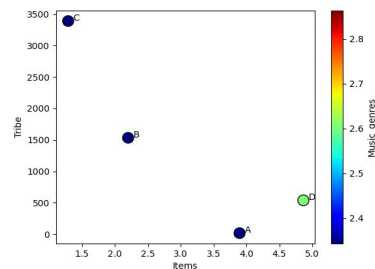
	Hit-Savvies (% over adopters)	leader Hit-Savvies (% over Hit-Savvies)	influenced Hit-Savvies (by leader Hit-Savvies)	neutral Hit-Savvies (% over Hit-Savvies)
Def1	2238 (1.21%)	489 (21%)	413 (183)	1436 (61,42%)
Def2	2307 (1.19%)	495 (21%)	421 (180)	1391 (60,29%)
Def3	2346 (1.21%)	493 (21%)	413 (183)	1437 (61,35%)
Def4	1864 (1.0%)	298 (16%)	249 (88)	1317 (70,65%)
Def5	1605 (0.96%)	328 (17%)	253 (93)	1352 (84,24%)
Def6	1072 (0.55%)	98 (8%)	75 (15)	905 (84,42%)
Def7	822 (0.42%)	86 (10%)	65 (12)	671 (81,63%)

Conclusioni

- è difficile quantificare il successo definito tramite *adoption trend* ;
- la purità dei Flop-adopters sensibilizza la predizione;
- è problematico identificare Hit-Savvies affidabili, che mantengano tale propensione nel tempo;
- la maggior parte degli Hit-Savvies non gode di potere d'influenza;
- l'influenza dei leaders, nella quasi totalità dei casi, deriva da relazioni di mutua amicizia;
- le relazioni di following/follower sono caratterizzate da un' assortatività topologica, nei gusti musicali, nel quantitativo e nei pattern di ascolto.



leaders as Hit-Savvies



**Grazie per
l'attenzione!**

