

The Netflix Dilemma: la percezione delle serie tv tra IMDB e Twitter

Laura Nembrini, Adele Zanfino, Antonio Lombardo, Gabriele Strano
Università degli Studi di Milano-Bicocca, CdLM Data Science

05 Giugno 2021

Sommario

Il mondo dell'intrattenimento ha subito profondi cambiamenti per via della comparsa di numerose piattaforme di streaming. Queste piattaforme sono riuscite inaspettatamente a surclassare la vendita di supporti fisici; per via della grande facilità di utilizzo, nonché della vasta scelta che si può avere a portata di click. Tra le numerose piattaforme esistenti, abbiamo scelto di analizzare Netflix, essendo quella più in voga negli ultimi anni. I contenuti di Netflix vengono periodicamente aggiornati (le nuove uscite mensili generano grande attesa tra i fruitori del servizio), generando quindi un catalogo in continuo cambiamento. Esiste un sito, chiamato IMDB, che si occupa di raccogliere, oltre alle informazioni sui vari film e le varie serie tv, anche le recensioni degli utenti, che vanno a comporre un indice di gradimento dei contenuti. Questo studio si propone dunque di utilizzare le informazioni raccolte, compresa un'analisi dei sentiment raccolti sul social network Twitter, per andare a capire quanto il rating fornito da IMDB sia veritiero rispetto al giudizio popolare sulle serie tv. Inoltre, si andrà a studiare il paese di produzione per capire quali fra questi ha un indice di gradimento maggiore.

Indice

| | | |
|----------|--|---|
| 1 | Introduzione | |
| 2 | Descrizione del dataset | |
| 3 | Obiettivo e target di riferimento | |
| 4 | Realizzazione delle infografiche | |
| 4.1 | Prima infografica | 3 |
| 4.2 | Seconda infografica | 3 |
| 4.3 | Terza Infografica | 3 |
| 5 | Valutazione delle infografiche | |
| 5.1 | Valutazione euristica | 4 |
| 5.2 | Test Psicometrico | 5 |
| 5.3 | Test Utente | 6 |

1 Introduzione

| | |
|----------|---|
| 1 | Per questo studio, abbiamo analizzato i database di Netflix, del sito IMDB e del social network |
| 2 | Twitter. Le caratteristiche estrapolate dai suddetti database, per creare il database finale integrato, sono elencate nel successivo paragrafo. |
| 2 | Inizialmente vediamo una dashboard dove compaiono tre riquadri, ognuno dei quali porta ad una diversa infografica. Nella prima, abbiamo proceduto nel mostrare la correlazione tra la durata del contenuto Netflix (espressa in numero di stagioni) e la media suo gradimento; nel secondo, la correlazione tra il paese di provenienza del contenuto e la popolarità (espressa in numero di tweet che si riferiscono alle serie prodotte |

nel suddetto paese); infine, nel terzo grafico è stata mostrata la correlazione tra la media del gradimento e la media del rating sul sito IMDB.

2 Descrizione del dataset

Per realizzare le infografiche, ci siamo avvalsi di un dataset ottenuto dall'integrazione dei tre dataset provenienti rispettivamente da Netflix, IMDB e Twitter. Il dataset finale è così composto:

- *Show_id*: il numero identificativo del contenuto;
- *Type*: indica se il contenuto è di tipo film o serie televisiva;
- *Title*: il titolo del contenuto;
- *Director*: il regista del contenuto;
- *Cast*: l'insieme degli attori più significativi;
- *Country*: il paese in cui è stato prodotto il contenuto;
- *Rating*: la classificazione secondo fasce d'età del contenuto;
- *Duration*: la durata del contenuto espressa come numero di stagioni;
- *Certificate*: certificato guadagnato da quella serie;
- *Runtime of episodes*: la durata in minuti dell'episodio;
- *Genre*: il genere artistico del contenuto;
- *IMBD_rating*: Il rating sul sito IMDB;
- *No_of_votes*: numero di voti degli utenti per ogni contenuto;
- *_id*: numero identificativo del tweet;
- *Text*: il testo del tweet;
- *Hashtags*: tag utilizzato su alcuni servizi web e social network come aggregatore tematico;

- *Score1*: colonna contenente un dizionario al cui interno vi sono le chiavi: positive, negative, neutral e compound ed i rispettivi valori;
- *Gradimento*: valore ricavato dalla sentiment analysis;
- *Text_cleaned*: il testo ripulito da eventuali emoticon, segni di punteggiatura e caratteri ininfluenti per la nostra indagine;

In totale, abbiamo analizzato i dati per 164 serie tv tra le più popolari su Netflix.

3 Obiettivo e target di riferimento

Il nostro obiettivo in questo studio è quello di mostrare quali caratteristiche dei contenuti Netflix potrebbero essere correlate con un maggior gradimento da parte del pubblico.

Il target a cui si riferisce questo studio si vuole composto di persone interessate a gettare un occhio sul come le serie tv sembrano essere percepite dal pubblico e sui fattori ipoteticamente correlati a questa percezione. Ciò è stato deciso per via del fatto che il numero dei fans di questi contenuti sta diventando progressivamente più alto, rendendo le serie tv uno degli argomenti più discussi nel web. I fruitori della piattaforma Netflix, e quindi il potenziale pubblico interessato all'argomento trattato, sono eterogenei, pertanto l'infografica è stata realizzata in modo semplice ed intuitivo, tale che non risultino necessarie particolari competenze per comprenderla.

4 Realizzazione delle infografiche

In accordo con le domande di ricerca prefissate, sono state realizzate delle infografiche che si propongono di rappresentare il gradimento delle serie tv presenti sulla piattaforma Netflix. Esse sono state realizzate sul software Tableau.

Qui il link alla dashboard iniziale:

<https://public.tableau.com/app/profile/adele.zanfino/viz/Schermatainiziale/SchermataIniziale>

4.1 Prima infografica

Come prima infografica, abbiamo realizzato uno scatterplot in cui si mostra la correlazione tra il numero di stagioni di una serie tv e il suo gradimento, tentando quindi di rispondere alla domanda:

“Quanto potrebbero influire la lunghezza di una serie sul suo gradimento?”

Qui di seguito viene mostrato il grafico:

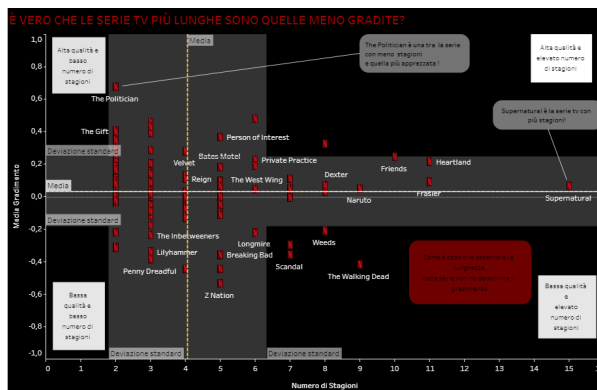


Figura 1: Scatterplot stagioni vs indice di gradimento.

Sull'asse delle ascisse abbiamo il numero delle stagioni di cui sono composte le serie prese in esame, su quello delle ordinate il gradimento. Quest'ultimo è misurato tramite l'analisi del Sentiment ottenuta dal social network Twitter e oscilla da un minimo di -1 ad un massimo di 1. Abbiamo aggiunto le etichette con i nomi delle varie serie tv in modo da poter capire immediatamente a cosa ci si riferiva. Com'è possibile osservare, la serie con maggior gradimento è "The politician", con 2 stagioni mentre quella con minor gradimento è "Z nation", con durata 5 stagioni. La serie tv "Supernatural", invece, è quella più lunga (15 stagioni), ma ha comunque un indice di gradimento al di sopra della media. Abbiamo indicato delle fasce di normalità basate sulla deviazione standard. Esso è un indice di variazione statistica, ovvero fornisce una stima della variabilità di una popolazione di dati. Nelle fasce prese in esame, può considerarsi contenuto circa il 68% delle osservazioni appartenenti a quella distribuzione, di fatti la gran parte delle serie tv considerate sono all'interno. Link per visualizzare l'infografica in modo inte-

rattivo:

<https://public.tableau.com/app/profile/adele.zanfino/viz/Gradimentovsstagioni/GradimentovsStagioni>

4.2 Seconda infografica

La seconda infografica ci mostra la correlazione tra la nazione di produzione (asse x) delle serie tv e la loro popolarità su Twitter (asse y), basandosi per quest'ultima grandezza sul conteggio dei tweet relativi ai contenuti presi in esame. Qui di seguito il grafico:

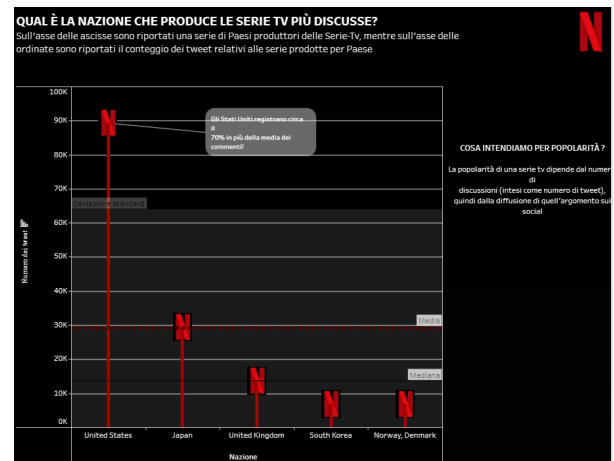


Figura 2: Lollipop nazioni vs numero di tweet.

Il tipo di grafico scelto in questo caso è di tipo lollipop, che rende facile mostrare la grande popolarità delle serie statunitensi rispetto a quelle di altri paesi. Le serie provenienti dagli USA totalizzano un ammontare di tweet quasi triplo rispetto alla seconda in classifica, il Giappone.

Link per visualizzare l'infografica in modo interattivo:

<https://public.tableau.com/app/profile/adele.zanfino/viz/Popolaritvspaese/PopolaritvsPaese>

4.3 Terza Infografica

Per finire, nella terza infografica da noi realizzata possiamo osservare la correlazione tra il rating IMDB (sull'asse delle ordinate) e le reazioni generate su Twitter (asse delle ascisse) riguardo alle serie tv prese in esame.

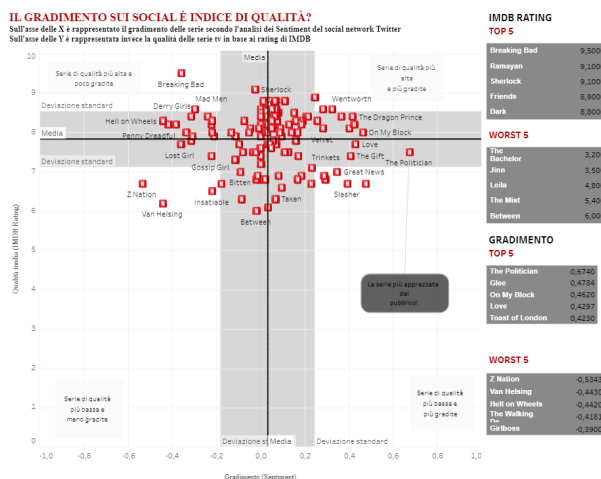


Figura 3: Scatterplot indice di gradimento vs rating IMDB.

Anche in questo caso abbiamo optato per un grafico scatterplot, con il rating espresso in numeri da 0 a 10 e il gradimento sui social riportato in numeri da -1 (per niente gradito) a 1 (molto gradito). Riteniamo interessante osservare che, mentre la media dei rating su IMDB si attesta poco al di sotto del voto 8, la media per il gradimento sui social staziona sullo 0. Infatti non sempre questi due modi di informarsi sul gradimento delle serie tv combaciano: possiamo prendere ad esempio la serie tv "Breaking Bad", che pur essendo prima nella classifica di IMDB, rimane al di sotto della media nell'analisi dei sentiment di Twitter.

Link relativo alla terza infografica:

<https://public.tableau.com/app/profile/adele.zanfino/viz/GradimentovsPopolarit/GradimentovsPopolarita>

5 Valutazione delle infografiche

Vi sono diversi modi per valutare un'infografica. Possiamo usare una valutazione assoluta (quanto un'infografica è buona) vs comparativa (quanto lo è rispetto ad un'altra infografica), formativa (compiuta durante il lavoro, per individuare e correggere eventuali errori in corso d'opera) vs sommativa (valutazione del risultato finale), e ovviamente qualitativa vs quantitativa.

Abbiamo tre principali tecniche di valutazione, di cui possiamo servirci per perfezionare i nostri grafici:

- **Valutazione euristica:** consiste nel chiedere ad un numero ristretto di persone (solitamente 6) una valutazione spontanea, a voce, del grafico. Essa ci consente di aumentare l'immediatezza del grafico, basandoci sugli eventuali problemi di comprensione avuti dalle persone. Come si può evincere, è un metodo di valutazione qualitativo;
- **Test psicometrico:** questo è un metodo che possiede sia caratteristiche quantitative che qualitative. Consiste nel sottoporre un questionario di gradimenti riguardante l'infografica, dando modo a chi sta valutando di adoperare una scala Likert per esprimere il proprio gradimento da 1 (molto basso) a 6 (molto alto). Ovviamente, il numero massimo e quello minimo possono essere impostati a seconda delle esigenze: un numero troppo basso di scelte potrebbe fornire una valutazione troppo sommaria, mentre uno troppo alto potrebbe confondere il soggetto, noi abbiamo deciso di utilizzare una scala da 1 a 6. Le caratteristiche che bisogna far valutare (o per lo meno le più importanti) sono: utilità, chiarezza, informatività, bellezza, intuitività e infine il valore complessivo. Questo test è da sottoporre ad almeno 24 utenti;
- **Test utente:** un tipo di test quantitativo, che serve a valutare efficacia ed efficienza delle nostre infografiche. Generalmente si somministra ad una quantità relativamente ristretta di persone (12 circa), e si sottopongono task molto specifici che devono portare a risposte ben definite. La stima dell'efficacia è data dal tasso d'errore mentre la valutazione dell'efficienza è data dal tempo medio di risposta.

5.1 Valutazione euristica

In questa valutazione abbiamo chiesto ad alcune persone di esprimere a voce alta ciò che pensavano delle infografiche mentre interagivano liberamente con esse.

- **Prima infografica:** abbiamo riscontrato che alcuni utenti non hanno riconosciuto

immediatamente i valori negativi sull'asse y. Abbiamo quindi proceduto ad inserire informazioni aggiuntive in forma scritta all'interno del grafico. Tuttavia, è stata particolarmente apprezzata l'idea di utilizzare il logo di Netflix per segnare i valori sul grafico;

- **Seconda Infografica:** una prima versione dell'infografica era stata fatta tramite un'infografica per bolle affiancate, in cui veniva proposto un confronto fra le serie più discusse sul web e le serie più gradite. Questa infografica non è stata molto apprezzata dagli utenti, alcune bolle erano infatti molto piccole e il grafico risultava di difficile comprensione. Abbiamo quindi provveduto a rifare completamente il grafico, secondo i suggerimenti del nostro pubblico. Per quanto riguarda la seconda versione, non sono stati rilevati problemi degni di nota secondo la valutazione euristica;
- **Terza infografica:** Nella valutazione della terza infografica, un utente ha fatto presente che condire di troppe vignette il grafico potrebbe privare chi lo adopera del piacere di scoprire autonomamente le particolarità dei dati contenuti. Abbiamo quindi eliminato le vignette superflue. Un altro utente ha fatto presente che presentare sull'asse x il valore della "popolarità" era errato, in quanto l'analisi dei sentiment di Twitter non compongono la popolarità ma il gradimento dei contenuti, o meglio ciò che la gente dice sui social riguardo a quel contenuto. Come ci è stato fatto notare, la popolarità è invece il valore espresso dall'asse delle y non in questa infografica ma in quella precedente. Abbiamo quindi proceduto a modificare il nome del valore sull'asse delle ascisse.

5.2 Test Psicometrico

Dopo aver modificato opportunamente le infografiche, successivamente alle valutazioni euristiche, abbiamo proceduto a somministrare un questionario psicometrico in cui è stato chiesto

di valutare le infografiche secondo la scala di Cabitza-Locoro. Essa permette di valutare la qualità delle visualizzazioni su una scala Likert da 1 a 6 in base ai seguenti aggettivi:

- Utile
- Chiara
- Informativa
- Bella
- Intuitiva
- Valore complessivo dell'infografica.

Al questionario hanno preso parte 37 persone. Qui di seguito mostriamo i Divergent stacked bar chart per ogni infografica in modo da poter visualizzare in maniera più chiara i risultati ottenuti.

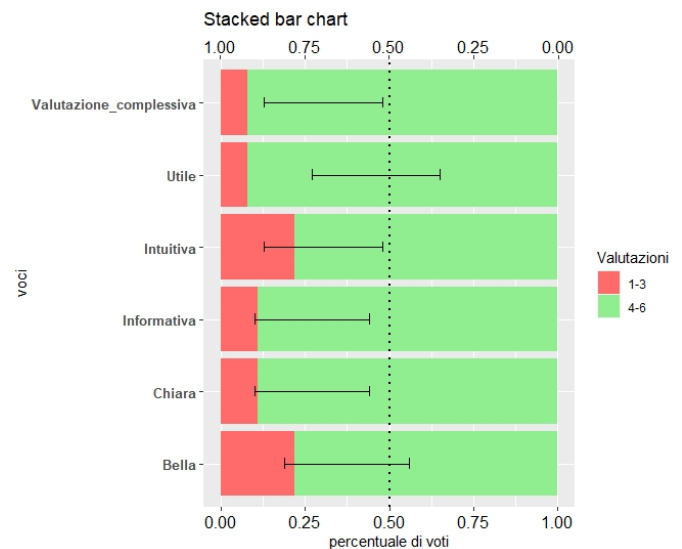


Figura 4: Test psicometrico prima infografica.

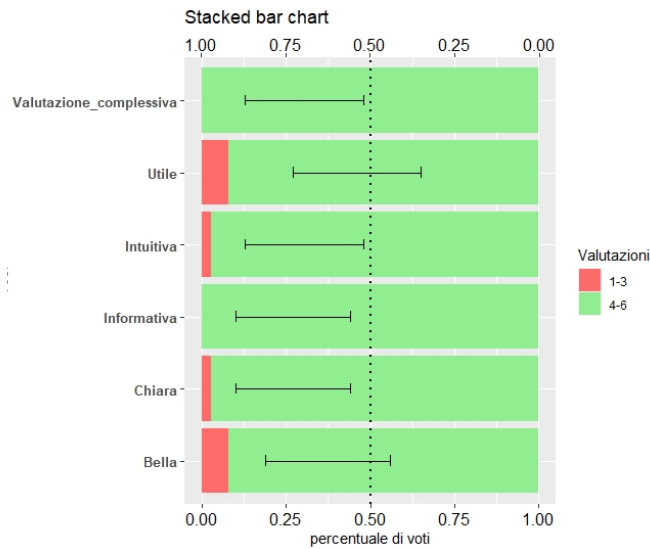


Figura 5: Test psicometrico seconda infografica.

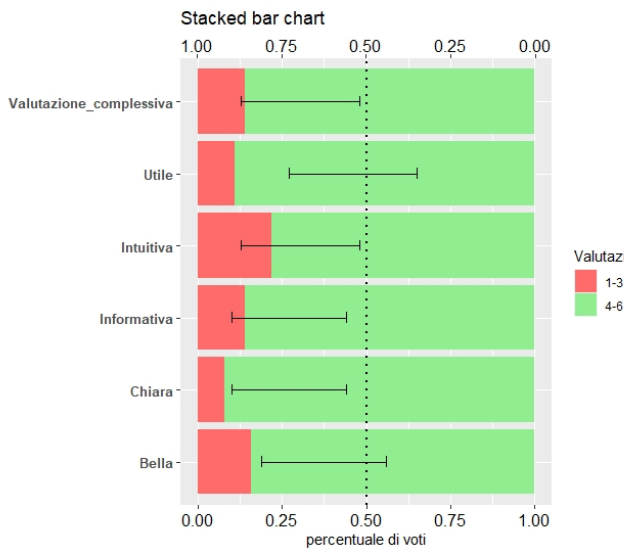


Figura 6: Test psicometrico terza infografica.

Nella seguente visualizzazione, abbiamo verificato se ci fosse correlazione tra le risposte date dagli utenti nei questionari.

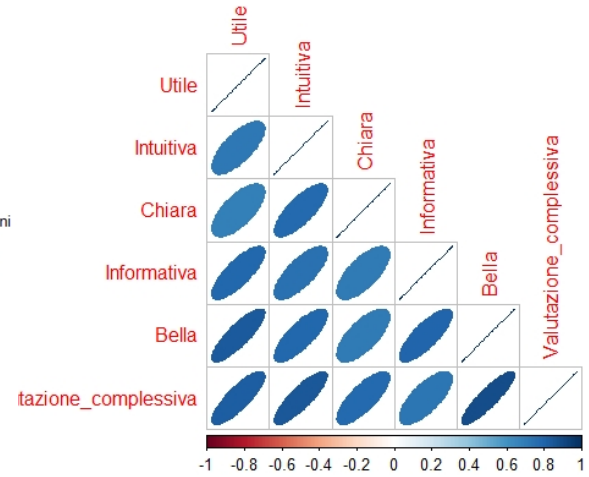


Figura 7: Correlazione.

Gli indici di correlazione più alti risultano in concomitanza delle seguenti coppie:

- Informativa-Utile;
- Bella-Utile;
- Valutazione complessiva-bella.

5.3 Test Utente

Come ultima valutazione è stato sottoposto il test utente a 15 soggetti. Il test utente consiste nel porre dei task, ossia delle domande, a cui gli utenti devono rispondere osservando le infografiche. Lo scopo è quello di capire se le informazioni contenute nelle infografiche siano percepite in modo corretto dai soggetti. Durante il test, per valutare l'efficienza, abbiamo misurato il tempo di esecuzione dei task per ciascun utente, mentre per quanto riguarda l'efficacia, questa è stata misurata considerando il numero dei task portati a termine correttamente. Per svolgere il test utente abbiamo individuato i seguenti 3 task:

- Per la prima infografica: quale tra le serie più lunghe è quella meno gradita dal pubblico?
- Per la seconda infografica: interagendo con il grafico, in quale paese le serie sono state più gradite?

- Per la terza infografica: considerando il rating IMDB, qual è la serie tv con indice di rating minore? Osservando le tabelle, quanto vale l'indice di rating medio per quella serie?

Riportiamo nelle tabelle sottostanti i risultati dei test per ciascun task. Nello specifico, nella colonna risultato, C sta ad indicare che l'utente ha inserito la risposta corretta mentre A la risposta errata.

| Utente | Tempo (s) | Risultato |
|--------|-----------|-----------|
| 1 | 33 | C |
| 2 | 20 | A |
| 3 | 40 | C |
| 4 | 6 | C |
| 5 | 33 | C |
| 6 | 16 | C |
| 7 | 28 | C |
| 8 | 25 | C |
| 9 | 40 | A |
| 10 | 20 | A |
| 11 | 60 | A |
| 12 | 70 | C |

Tabella 1: Risultati task 1.

| Utente | Tempo (s) | Risultato |
|--------|-----------|-----------|
| 1 | 22 | C |
| 2 | 5 | C |
| 3 | 30 | C |
| 4 | 11 | A |
| 5 | 13 | C |
| 6 | 24 | C |
| 7 | 15 | C |
| 8 | 10 | A |
| 9 | 35 | C |
| 10 | 40 | C |
| 11 | 20 | C |
| 12 | 40 | C |

Tabella 2: Risultati task 2.

Di seguito riportiamo inoltre graficamente gli esiti dello svolgimento dei tre task con i relativi intervalli di errore standard.

| Utente | Tempo (s) | Risultato |
|--------|-----------|-----------|
| 1 | 139 | C |
| 2 | 15 | C |
| 3 | 40 | A |
| 4 | 16 | C |
| 5 | 45 | C |
| 6 | 35 | C |
| 7 | 13 | C |
| 8 | 15 | A |
| 9 | 40 | A |
| 10 | 50 | A |
| 11 | 40 | A |
| 12 | 30 | A |

Tabella 3: Risultati task 3.

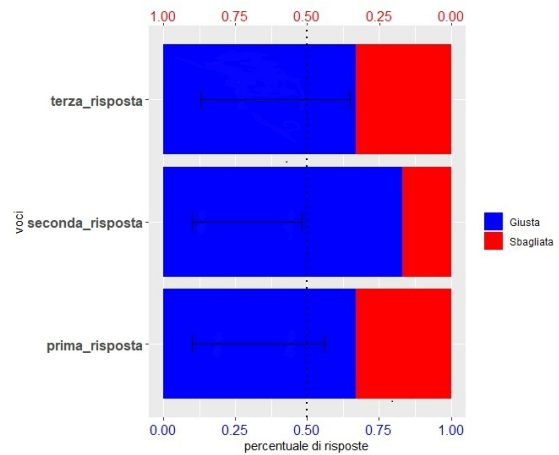


Figura 8: Risultati test utente.

Nella seconda infografica non abbiamo riscontrato particolari errori, al contrario della prima e della terza in quanto leggermente più complicate della seconda e per le quali veniva richiesto uno sforzo di comprensione maggiore.

Per quanto riguarda i tempi di esecuzione invece riportiamo i seguenti violin plot per ciascun task.

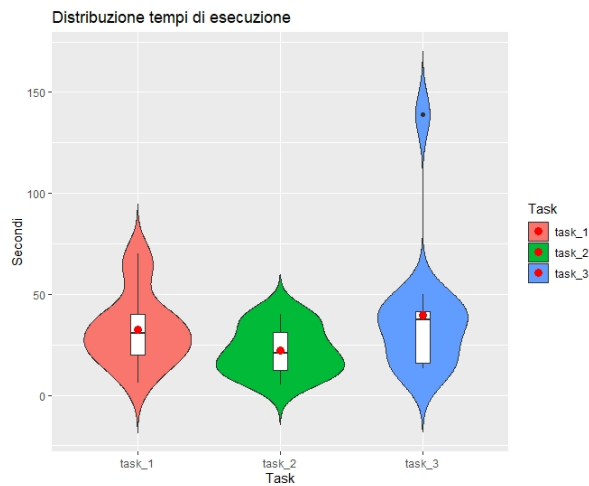


Figura 9: Tempi di esecuzione per task.

Dalla Figura 9 possiamo ricavare i seguenti indici per ciascun task:

- Task 1:
 - Media: 32.58 s
 - Deviazione Standard: 18.18 s
 - Mediana: 30.5 s
- Task 2:
 - Media: 22.08 s
 - Deviazione Standard: 11.96 s
 - Mediana: 21 s
- Task 3:
 - Media: 39.83 s
 - Deviazione Standard: 33.85 s
 - Mediana: 37.5 s

Come potevamo aspettarci, la media dei tempi di esecuzione associata al task 3 è maggiore rispetto alle altre in quanto quello specifico task richiedeva un maggior sforzo interpretativo. Si può inoltre notare, sia dal grafico che dai valori riportati, che il task 3 presenta una maggiore variabilità con la presenza anche di un unico outlier con tempo di esecuzione pari circa a 140.