



Manuale Utente

Gruppo QuaranTeam - Progetto HD Viz

quaranteam2021@gmail.com

INFORMAZIONI SUL DOCUMENTO

Versione	2.0.0
Approvatore	Consalvo Federico
Redattori	Veronese Luca Mason Damiano
Verificatori	Sinigaglia Matteo Rech Elia
Stato	Approvato
Uso	Esterno
Destinato a	Prof. Vardanega Tullio Prof. Cardin Riccardo Zucchetti S.p.A. QuaranTeam

Descrizione

Manuale utente per l'installazione e l'utilizzo di *HD Viz* realizzato dal gruppo *QuaranTeam*.

Registro delle modifiche

Versione	Data	Descrizione	Autore	Verificatore
2.0.0	2021-06-10	Approvazione del documento.	Consalvo Federico	
1.0.2	2021-06-07	Modifica §4.	Mason Damiano	Sinigaglia Matteo
1.0.1	2021-06-06	Modifica §3.	Veronese Luca	Rech Elia
1.0.0	2021-05-10	Approvazione del documento.	Veronese Luca	
0.6.0	2021-05-04	Stesura §A.	Gibellato Alice, Sinigaglia Matteo	Chiarello Federico
0.5.0	2021-05-03	Stesura §4.	Sinigaglia Matteo	Chiarello Federico
0.4.0	2021-05-02	Stesura §3.	Sinigaglia Matteo	Mason Damiano
0.3.0	2021-04-28	Stesura §2.	Gibellato Alice	Chiarello Federico
0.2.0	2021-04-24	Stesura §1.	Gibellato Alice	Mason Damiano
0.1.0	2021-04-24	Creazione del documento.	Veronese Luca	Chiarello Federico

Indice

1	Introduzione	1
1.1	Scopo del documento	1
1.2	Scopo del prodotto	1
1.3	Glossario	1
1.4	Riferimenti	1
1.5	Supporto tecnico	2
2	Installazione	3
2.1	Requisiti	3
2.2	Installazione e avvio	3
2.2.1	Preliminari	3
2.2.2	Avvio del client	4
2.2.3	Configurazione ed avvio del server (opzionale)	4
3	Guida all'utilizzo	6
3.1	Navigazione nel sito	6
3.2	Operazioni sui grafici	6
3.2.1	Inserimento	6
3.2.2	Modifiche al grafico	8
3.2.2.1	Modifica titolo	8
3.2.2.2	Interazioni con il grafico	8
3.2.3	Rimozione di un grafico	9
3.3	Visualizzazione manuale e guida utente	10
3.4	Operazioni sul database	11
3.4.1	Importazione dati	11
3.4.2	Aggiunta di un dataset	11
3.4.3	Eliminazione di un dataset	12
4	Componenti e Funzionalità	13
4.1	Caricamento dati	13
4.2	Scelta di una riduzione dimensionale _G	14
4.3	Scelta di una funzione per il calcolo delle distanze	15
4.4	Selezione della normalizzazione	15
4.5	Funzionalità dei grafici	16
4.5.1	Force Field	16
4.5.1.1	Modalità di interazione	16
4.5.1.2	Parametri della visualizzazione	16
4.5.2	Heat Map	16
4.5.2.1	Modalità di interazione	17
4.5.2.2	Parametri della visualizzazione	17
4.5.3	Scatter plot Matrix	18
4.5.3.1	Modalità di interazione	18
4.5.4	Proiezione Lineare Multi Asse	19
4.5.4.1	Modalità di interazione	19
4.6	Complementi ai grafici	20

4.6.1	Legenda	20
A	Glossario	21

Elenco delle figure

1	Schermata principale	6
2	Inserimento di un grafico	7
3	Modifica al nome di un grafico	8
4	Modifiche ai grafici	9
5	Visualizzazione Manuale Utente	10
6	Visualizzazione Guida Utente	10
7	Importazione di un dataset dal database	11
8	Operazioni con il database	12
9	Modifica parametri Force Field	17
10	Modifica parametri Heat Map	18
11	Interazione con Scatter plot Matrix	19
12	Interazione con Proiezione Lineare Multi Asse	20

Elenco delle tabelle

2	Manipolazioni possibili per ogni grafico	8
3	Interazioni possibili per ogni grafico	9
4	Opzioni disponibili per ogni riduzione dimensionale	14

1 Introduzione

1.1 Scopo del documento

Il presente documento rappresenta il manuale utente per l'applicazione web *HD Viz*. Vengono descritte dettagliatamente tutte le caratteristiche dell'applicativo utilizzabili dall'utente.

Tale manuale è suddiviso in sezioni:

- in §2 vengono riportate le istruzioni per l'installazione;
- in §3 vengono descritte le operazioni principali eseguibili dall'utente di *HD Viz*;
- in §4 il focus è rivolto verso le componenti visibili dell'applicazione, quindi vengono descritte le loro caratteristiche, le loro funzionalità e le loro modalità di interazione.

Il manuale sarà consultabile alla sezione "Aiuto" accessibile dal menu della pagina principale di *HD Viz*.

1.2 Scopo del prodotto

HD Viz è un'applicazione web avente scopo di fornire uno strumento per la visualizzazione di dati con molte dimensioni a supporto della fase esplorativa dell'analisi dei dati. *HD Viz* è in grado di rappresentare dati che possono avere almeno 15 dimensioni e fornisce 4 diversi tipi di visualizzazione a tale scopo.

1.3 Glossario

Il documento contiene termini che possono presentare significati ambigui. Viene quindi fornito un glossario individuabile nell'appendice §A, all'interno del documento, contenente tutti i termini definiti ambigui e la loro spiegazione. Nel documento vengono identificati con una G a pedice.

1.4 Riferimenti

Per una trattazione più approfondita di funzioni matematiche, algoritmi e grafici sono state selezionate le seguenti fonti:

- **Funzioni di riduzione dimensionale_G:**
 - FASTMAP:
<https://dl.acm.org/doi/10.1145/568271.223812>
 - ISOMAP:
<https://science.sciencemag.org/content/290/5500/2319>
 - t-SNE:
<https://www.jmlr.org/papers/v9/vandermaaten08a.html>
 - LLE:
<https://science.sciencemag.org/content/290/5500/2323>
 - UMAP:
<https://arxiv.org/abs/1802.03426>
- **Funzioni di calcolo delle distanze:**

- Distanza_G di Canberra:
https://en.wikipedia.org/wiki/Canberra_distance
 - Distanza_G Euclidea e Euclidea Quadrata:
<https://xlinux.nist.gov/dads/HTML/euclidndstnc.html>
 - Distanza_G di Chebyshev:
https://en.wikipedia.org/wiki/Chebyshev_distance
 - Distanza_G Cosine:
<https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/cosdist.htm>
 - Distanza_G di Manhattan:
<https://xlinux.nist.gov/dads/HTML/manhattanDistance.html>
- **Grafici:**
 - Force Field:
<https://observablehq.com/@d3/force-directed-graph>
 - Heat Map:
<https://observablehq.com/@bstaats/matrix-diagram>
 - Scatter plot Matrix:
<https://observablehq.com/@d3/scatterplot-matrix>
 - Proiezione Lineare Multi Asse:
<https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/visualize/freeviz.html>

1.5 Supporto tecnico

Qualora venissero riscontrati bug_G o malfunzionamenti, si prega di segnalarli per mail all'indirizzo:

quaranteam2021@gmail.com

È preferibile l'utilizzo della seguente struttura:

Oggetto: HD Viz - bug - < Evento da segnalare >;

Allegato (opzionale): è possibile allegare immagini o video esplicativi;

Corpo:

- browser_G e sistema operativo utilizzati;
- descrizione del problema;
- versione di HD Viz utilizzata.

2 Installazione

2.1 Requisiti

Sistema operativo

- Windows 10;
- Ubuntu 20.04 o superiore, o distribuzioni Linux derivate.

Browser

- **Mozilla Firefox** v85.0 o superiore;
- **Google Chrome** v87.0 o superiore.

È consigliabile avere impostato uno dei due browser_G elencati come browser_G predefinito, perché la procedura indicata in §2.2.2 (Installazione - Avvio del client) apre in automatico una pagina del browser_G predefinito. È comunque possibile aprire l'applicazione da un browser_G non predefinito in seguito, al seguente URL:

`http:localhost/<numero_porta>`

Non è garantita la piena compatibilità per browser_G derivati dai suddetti o per versioni precedenti dei suddetti.

Requisiti software

- **Node.js_G** 14.16.0 o superiore;
- **npm_G** 7.6.3 o superiore;
- **PostgreSQL_G** 13.2 o superiore (opzionale¹).

Requisiti hardware

- **RAM:** necessari almeno 4 GB, consigliati almeno 8 GB;
- **connessione a Internet.**

2.2 Installazione e avvio

2.2.1 Preliminari

1. Aprire la pagina del repository_G GitHub_G di *HD Viz* al seguente link:

<https://github.com/QuaranTeam2021/HD-Viz>

2. Selezionare **Code > Download Zip**;
3. Salvare l'archivio *.zip* ed estrarre i file nella locazione desiderata;
4. Aprire la cartella *HD-Viz-main*.

¹necessario solo per il collegamento con il database_G

In alternativa, in un sistema con `wget` e `unzip` installati² aprire il terminale³ e digitare le seguenti istruzioni:

```
wget https://github.com/QuaranTeam2021/HD-Viz/archive/main.zip
unzip main.zip
cd HD-Viz-main
```

2.2.2 Avvio del client

Una volta posizionati all'interno della cartella `HD-Viz-main`:

1. Aprire la cartella `client` e aprire un terminale posizionato nella cartella `client` (se non già aperto);
2. Se si cerca di avviare per la prima volta *HD Viz*, digitare il comando:

```
npm install
```

3. Eseguire la `buildG` dell'applicazione con il comando:

```
npm run build
```

4. Se si cerca di avviare per la prima volta *HD Viz*, digitare il comando:

```
npm install -g serve
```

5. Digitare infine il comando:

```
serve -s build -l 3000
```

6. Attendere l'apertura automatica del browser_G predefinito.

2.2.3 Configurazione ed avvio del server (opzionale)

Il server_G viene sfruttato per interagire con un database_G esterno, pertanto la sua installazione ed esecuzione è opzionale. Il database_G utilizzato è PostgreSQL, scaricabile al seguente link:

<https://www.postgresql.org/download/>

Una volta posizionati all'interno della cartella `HD-Viz-main`:

1. Se si cerca di avviare per la prima volta la componente server, aprire un terminale nella cartella `HD-Viz-main` e digitare il comando:

```
npm install
```

2. Aprire la cartella `server` e creare un file `.env` rispettando la seguente notazione:

²installati di default in ogni distribuzione Linux, necessaria installazione manuale per Windows

³terminale per Linux o Windows PowerShell per Windows

```
HDVIZ_USER = postgres
HDVIZ_PASSWORD = postgres
HDVIZ_HOST = localhost
HDVIZ_PORT = 5432
HDVIZ_DATABASE = hdviz
```

Inserire come valore dei campi HDVIZ_USER, HDVIZ_PASSWORD e HDVIZ_DATABASE rispettivamente il nome utente, la password e il nome del database_G a cui ci si vuole connettere.

3. Spostarsi con il terminale nella cartella `server` e avviare il server_G con il comando:

```
node server
```

Il server_G sarà in esecuzione alla porta 5000.

3 Guida all'utilizzo

3.1 Navigazione nel sito

All'apertura dell'applicazione viene visualizzata la schermata sottostante:



Figura 1: Schermata principale

Ogni pulsante svolge la funzione brevemente descritta sopra di esso. Una volta selezionata una pagina i pulsanti appaiono in alto nella nuova pagina, come mostrato in figura 2. È sempre possibile tornare alla pagina principale premendo il tasto "Home".

3.2 Operazioni sui grafici

3.2.1 Inserimento

È possibile importare un dataset dalla memoria locale o dal database_G, scegliendo tra i due pulsanti mostrati in figura 2.

- 1.a) Cliccare sul pulsante "IMPORTA" e scegliere il file .csv_G o .tsv_G da cui prelevare i dati. Per ulteriori informazioni su caratteristiche obbligatorie dei file .csv_G o .tsv_G si rimanda a §4.1.

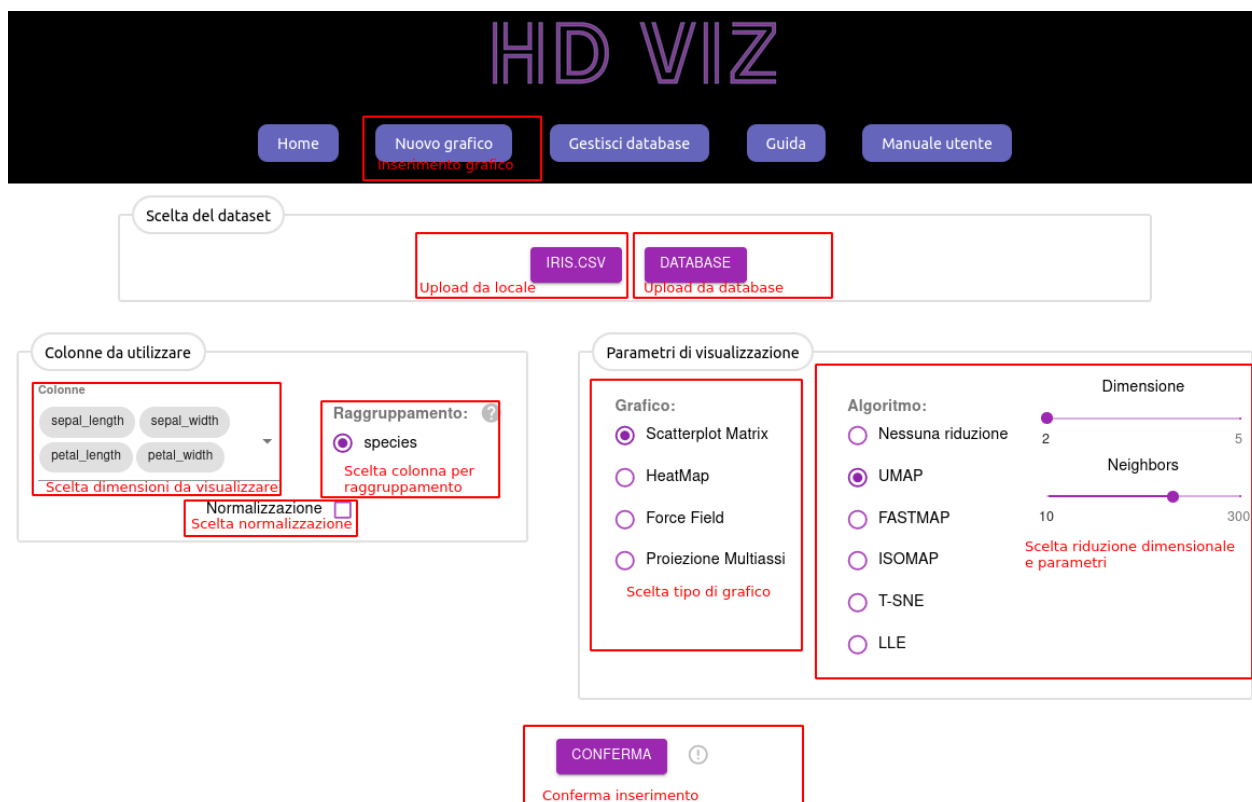
Oppure

- 1.b) Cliccare sul pulsante "DATABASE". Nella finestra che si aprirà saranno presenti i seguenti due menù a tendina:
 - **Tabelle:** in cui è possibile selezionare uno dei dataset presenti nel database_G;
 - **Colonne:** in cui è possibile selezionare le dimensioni del dataset selezionato che si vogliono importare.

Per ulteriori dettagli sull'importazione di dataset dal database_G si rimanda a §3.4.1.

Una volta importato un dataset, è possibile selezionare una visualizzazione ed eventualmente impostare le opzioni di manipolazione dati disponibili, attraverso l'interfaccia di figura 2:

- 2) Selezionare una o più colonne da visualizzare;
- 3) Selezionare una o nessuna colonna per il raggruppamento;
- 4) Selezionare una tra le possibili visualizzazioni;
- 5) Scegliere se applicare o meno la normalizzazione ai dati (si veda §4.4 - Normalizzazione);
- 6) Selezionare le opzioni per la manipolazione dati. Tali opzioni dipendono dal grafico scelto al punto 4. La gamma di opzioni presenti per ogni grafico è riassunta nella tabella 2.
- 7) Premere il pulsante "CONFERMA" per aggiungere il grafico alla pagina di visualizzazione ed essere rimandati a tale pagina.



The screenshot shows the HD VIZ interface for adding a new chart. The interface is divided into several sections:

- Sceita del dataset:** Contains two buttons: "IRIS.CSV" (labeled "Upload da locale") and "DATABASE" (labeled "Upload da database").
- Colonne da utilizzare:** Contains a list of columns: "sepal_length", "sepal_width", "petal_length", and "petal_width". Below this is a "Raggruppamento:" dropdown menu with "species" selected. There is also a "Normalizzazione" checkbox.
- Parametri di visualizzazione:** Contains a "Grafico:" list with options: "Scatterplot Matrix", "HeatMap", "Force Field", and "Proiezione Multiassi". It also contains an "Algoritmo:" list with options: "Nessuna riduzione", "UMAP", "FASTMAP", "ISOMAP", "T-SNE", and "LLE". There are sliders for "Dimensione" (ranging from 2 to 5) and "Neighbors" (ranging from 10 to 300).
- CONFERMA:** A button at the bottom labeled "CONFERMA" with a question mark icon, and a label "Conferma inserimento" below it.

Figura 2: Inserimento di un grafico

Tabella 2: Manipolazioni possibili per ogni grafico

	Scatter plot Matrix	Force Field	Heat Map	Proiezione Lineare Multi Asse
Normalizzazione di dati	Sì	Sì	Sì	Sì
Visualizzazione di una riduzione dimensionale	Sì	No	No	Sì
Visualizzazione di una matrice delle distanze	No	Sì	Sì	No
Visualizzazione di dati non manipolati	Sì	No	No	Sì

3.2.2 Modifiche al grafico

3.2.2.1 Modifica titolo

- 1) Aprire il pannello delle opzioni di un grafico;
- 2) Cliccare sul pulsante con l'icona di modifica (figura 3.a);
- 3) Inserire il nome desiderato per il grafico;
- 4) Premere l'icona di conferma (figura 3.b).

A questo punto il nuovo titolo viene visualizzato. È possibile modificare nuovamente il titolo.

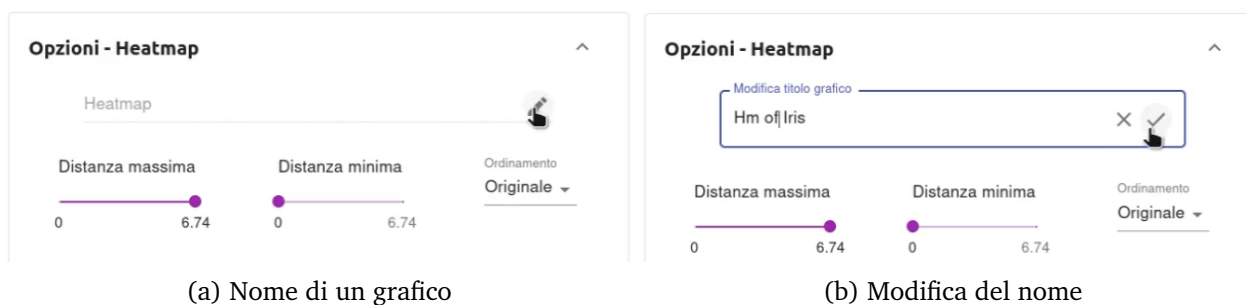


Figura 3: Modifica al nome di un grafico

3.2.2.2 Interazioni con il grafico

Uno dei punti di forza di *HD Viz* è la presenza di grafici *dinamici*, che rispondono alle interazioni con l'utente. Le opzioni dipendono dallo specifico tipo di grafico, pertanto si rimanda alla sezione §4.5 che illustra in dettaglio le caratteristiche e le funzionalità presenti per ogni grafico. Un riassunto delle interazioni possibili è il seguente:

Il procedimento per l'applicazione di modifiche è identico per tutti i grafici:

- 1) Aprire il pannello delle opzioni di un grafico;

Tabella 3: Interazioni possibili per ogni grafico

	Scatter plot Matrix	Force Field	Heat Map	Proiezione Lineare Multi Asse
Evidenziamento di dati	Sì	No	No	No
Aggiunta dimensione	Sì	No	No	Sì
Rimozione dimensione	Sì	No	No	Sì
Ordinamento elementi	No	No	Sì	No
Trascinamento nodo	No	Sì	No	No
Modifica intensità della forza applicata	No	Sì	No	No
Modifica intervallo di distanze visualizzate	No	Sì	Sì	No

2) Modificare una o più opzioni tra quelle disponibili;

3) Premere il pulsante di conferma.

3.2.3 Rimozione di un grafico

È possibile rimuovere un grafico dalla lista dei grafici visualizzati:

1) Aprire il pannello delle opzioni di un grafico;

2) Premere il pulsante "RIMUOVI".

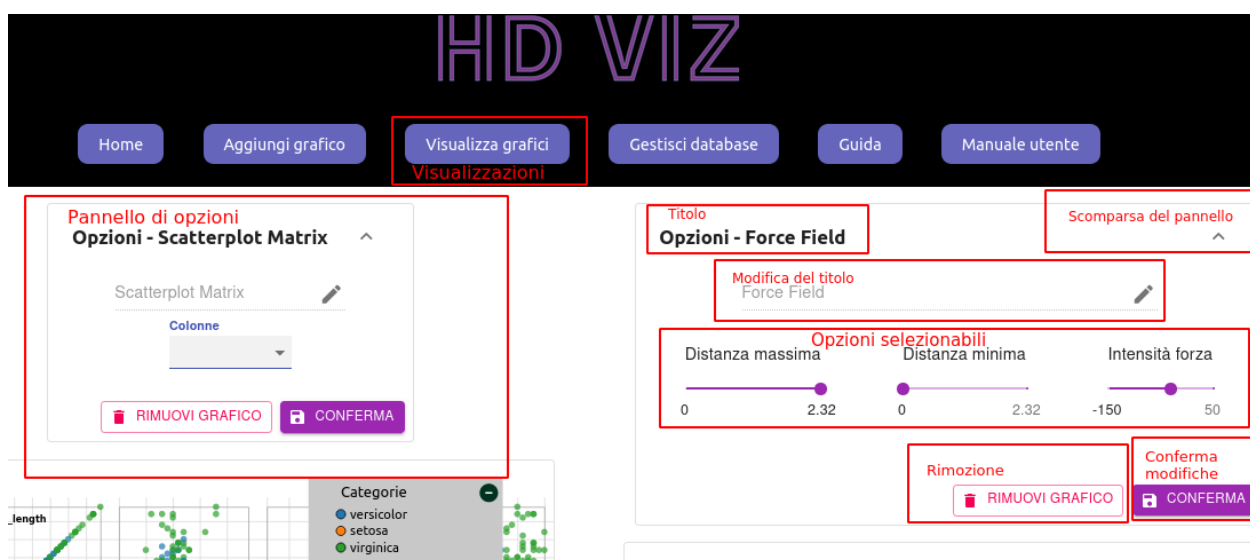


Figura 4: Modifiche ai grafici

3.3 Visualizzazione manuale e guida utente

Per visualizzare il *Manuale Utente* (il presente documento) selezionare dal menù in alto la voce "Manuale Utente".

Per visualizzare la Guida Utente selezionare dal menù principale la voce "Guida".



Figura 5: Visualizzazione Manuale Utente



Figura 6: Visualizzazione Guida Utente

3.4 Operazioni sul database

3.4.1 Importazione dati

Dopo aver cliccato sul pulsante "DATABASE" come modalità di importazione dei dati scelta, si aprirà una finestra con i seguenti due menù (figura 7):

- **Tabelle:** in cui è possibile selezionare uno dei dataset presenti nel database;
- **Colonne:** in cui è possibile selezionare le dimensioni del dataset selezionato che si vogliono importare.

Nel caso in cui non vengano selezionate delle colonne tramite il menù "Colonne", verrà importato l'intero dataset con tutte le dimensioni presenti nel database_G.

Al termine delle selezioni è sufficiente cliccare il pulsante "SCEGLI QUESTI DATI" per completare l'operazione di importazione.

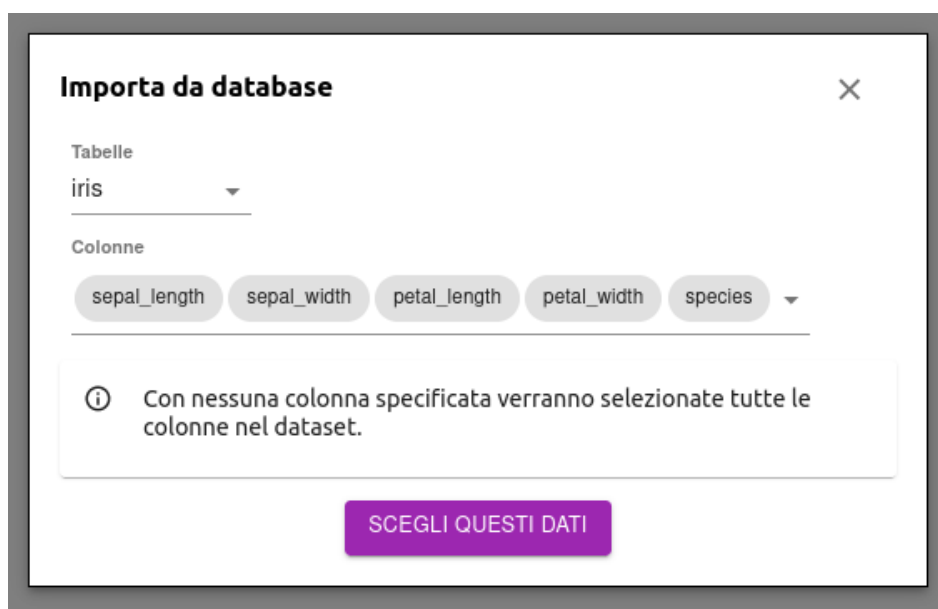


Figura 7: Importazione di un dataset dal database

3.4.2 Aggiunta di un dataset

Accedendo alla sezione "Gestisci Database" è possibile caricare i dati contenuti in un file CSV_G o TSV_G nel database_G esterno tramite i seguenti passaggi (figura 8):

1. Cliccare sul pulsante di aggiunta (+) e scegliere il file .csv_G o .tsv_G contenente i dati che vogliamo caricare nel database_G. Per ulteriori informazioni su caratteristiche obbligatorie dei file .csv_G o .tsv_G si rimanda a §4.1;
2. Inserire il nome con cui si vuole salvare il dataset nel database_G nell'area di testo apposita. Nel caso in cui non venga esplicitamente inserito un nome verrà utilizzato il nome del file caricato;
3. Confermare l'operazione premendo il tasto "INVIA".

3.4.3 Eliminazione di un dataset

Accedendo alla sezione "Gestisci Database" è possibile eliminare un dataset contenuto nel database, tramite i seguenti passaggi (figura 8) cliccando sul pulsante con l'icona di eliminazione situato a fianco del nome del dataset che vogliamo eliminare.

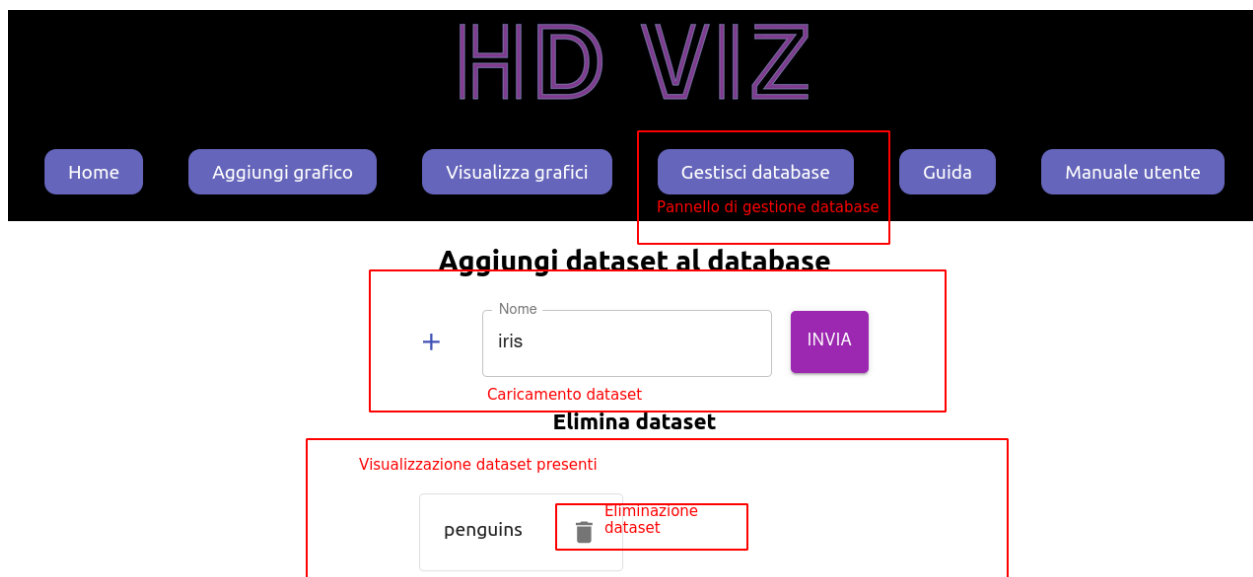


Figura 8: Operazioni con il database

4 Componenti e Funzionalità

4.1 Caricamento dati

L'utente preme sul pulsante importa file e seleziona un file con estensione `.csvG` o `.tsvG`. Questo file deve contenere *obbligatoriamente*:

- **una e una sola** riga di intestazione;
- **una o più** righe di dati.

Un esempio di file `.csvG` ben formato è il seguente:

Listing 1: formattazione corretta CSV

```
sepal_length,sepal_width,petal_length,petal_width,species
5.1,3.5,1.4,0.2,setosa
4.9,3.0,1.4,0.2,setosa
4.7,3.2,1.3,0.2,setosa
4.6,3.1,1.5,0.2,setosa
5.0,3.6,1.4,0.2,setosa
```

Un esempio di file `.tsvG` ben formato è il seguente:

Listing 2: formattazione corretta TSV

sepal_length	sepal_width	petal_length	petal_width	species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa

Il contenuto del file in input deve sottostare alle seguenti regole:

- il carattere delimitatore per un file `.csvG` deve essere una virgola ',' o un punto e virgola ';' ;
- il carattere delimitatore per un file `.tsvG` deve essere una tabulazione;
- il delimitatore per le cifre decimali deve essere un punto;
- non deve essere presente un delimitatore per raggruppare le migliaia. Ciò implica che:
 - 10000 è accettato e considerato come 10^4 ,
 - 10.000 è accettato e considerato come 10^1 ,
 - 10'000 non è accettato,
 - 10,000 viene interpretato come il contenuto di due celle contigue;
- ogni colonna deve contenere dati dello stesso tipo;
- il formato di codifica accettato è UTF-8.

Una volta importati i dati l'utente può selezionare tra le colonne numeriche quelle da considerare per la visualizzazione senza manipolazioni, per la riduzione dimensionale_G o per il calcolo della matrice delle distanze_G.

In seguito è possibile selezionare la colonna (numerica o non numerica) per il raggruppamento. È possibile validare ed eventualmente correggere un file `.csvG` all'indirizzo

<http://www.fixcsv.com/>

4.2 Scelta di una riduzione dimensionale_G

Dopo aver scelto una visualizzazione Scatter plot Matrix o Proiezione Lineare Multi Asse, è possibile selezionare **una sola o nessuna** delle seguenti riduzioni dimensionali_G e le relative opzioni:

- FASTMAP, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
- ISOMAP, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
 - parametro *neighbors*;
- t-SNE, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
 - parametro *neighbors*;
 - parametro *perplexity*;
 - parametro *epsilon*;
- LLE, con opzioni:
 - numero di dimensioni in output;
 - funzione di calcolo della distanza_G associata;
 - parametro *neighbors*;
- UMAP, con opzioni:
 - numero di dimensioni in output;
 - parametro *neighbors*.

Le opzioni possibili sono riassunte nella tabella in seguito:

Tabella 4: Opzioni disponibili per ogni riduzione dimensionale

	FASTMAP	ISOMAP	t-SNE	LLE	UMAP
Numero di dimensioni in output	Sì	Sì	Sì	Sì	Sì
Funzione di calcolo della distanza	Sì	Sì	Sì	Sì	No
Parametro <i>neighbors</i>	No	Sì	Sì	Sì	Sì
Parametro <i>perplexity</i>	No	No	Sì	No	No
Parametro <i>epsilon</i>	No	No	Sì	No	No

Le varie opzioni appaiono e scompaiono a video dinamicamente in seguito alla selezione della riduzione dimensionale_G. Tutte le opzioni sono selezionabili tramite slider_G numerico, ad eccezione della scelta della funzione di calcolo della distanza_G associata, selezionabile tramite lo stesso widget utilizzato per la scelta di una funzione per il calcolo delle distanze_G utilizzato per il calcolo della matrice delle distanze_G.

È consentito selezionare un numero di dimensioni in output minore, uguale o maggiore del numero di dimensioni in input. Alcune riduzioni dimensionali possono generare dei valori nulli in presenza di particolari configurazioni del dataset.

Per una trattazione approfondita dei vari algoritmi di riduzione dimensionale_G si rimanda a §1.4 - Riferimenti.

4.3 Scelta di una funzione per il calcolo delle distanze

Dopo aver selezionato una visualizzazione Force Field o Heat Map è necessario scegliere **una e una sola** tra le funzioni disponibili per il calcolo delle distanze_G.

Le opzioni possibili sono le seguenti:

- distanza_G Euclidea;
- distanza_G di Manhattan;
- distanza_G Cosine;
- distanza_G di Chebyshev;
- distanza_G Euclidea quadrata;
- distanza_G di Canberra.

Sono selezionabili tramite menu a tendina.

Il calcolo della distanza può produrre valori non numerici in presenza di alcune configurazioni del dataset (es. il calcolo della distanza di Canberra comporta una divisione con denominatore potenzialmente uguale a 0). In caso vengano prodotti valori non validi questi non vengono mostrati nella visualizzazione, e viene mostrato un avviso nella legenda del grafico contenente il numero di valori non numerici individuati.

Per una trattazione approfondita delle varie funzioni di calcolo della distanza_G si rimanda a §1.4 - Riferimenti

4.4 Selezione della normalizzazione

HD Viz permette di scegliere se applicare o meno l'operazione di *normalizzazione* sui dati, prima di applicare un'eventuale manipolazione su questi. L'operazione di *normalizzazione* individua e applica una funzione per ogni colonna del dataset, avente come dominio l'insieme dei valori della colonna del dataset e come codominio l'intervallo chiuso $[0, 1]$. Il valore più alto del dominio (che chiamiamo m) viene associato a 1, l'immagine (x_i) di ogni altro valore (x_d) del dominio assume il valore $(\frac{x_d}{m})$.

Tale operazione serve ad equiparare il contributo delle colonne del dataset nella visualizzazione, a prescindere dall'ampiezza del dominio in cui i valori sono distribuiti. La *normalizzazione* può essere desiderabile o meno.

4.5 Funzionalità dei grafici

4.5.1 Force Field

Il grafico Force Field (figura 9) permette di visualizzare un integratore numerico di velocità Verlet_G per la simulazione di forze fisiche sulle particelle, utile per lo studio di reti.

4.5.1.1 Modalità di interazione

È possibile *trascinare* e *fissare* il nodo in un punto dello schermo, **mantenendo premuto** il cursore sopra un nodo e **spostando** il nodo nel punto desiderato. Al termine dell'operazione il bordo del nodo assume il colore nero.

È possibile *rilasciare* un nodo precedentemente fissato, **cliccando** sul nodo stesso. Al termine dell'operazione il bordo del nodo assume il colore bianco.

È possibile *visualizzare informazioni* associate a un nodo, **posizionando il cursore** sopra il nodo stesso.

4.5.1.2 Parametri della visualizzazione

Attraverso il pannello di controllo della visualizzazione si possono modificare i seguenti parametri:

- **Intensità della forza_G**: l'utente può impostare l'attrazione reciproca esercitata dai nodi. Se impostata ad un valore positivo i nodi si attraggono tra loro, se impostata ad un valore negativo i nodi si respingono. Il suo valore di default è -30 , l'intervallo chiuso nel quale può spaziare è $[-150, 50]$.
- **Distanza_G massima**: distanza_G massima tra i nodi affinché l'arco che collega la coppia di nodi venga inserito nella simulazione di forza e venga visualizzato. Il suo valore di default è pari alla distanza massima tra coppie di nodi della grafo.
- **Distanza_G minima**: distanza_G minima tra i nodi affinché l'arco che collega la coppia di nodi venga inserito nella simulazione di forza e venga visualizzato. Il suo valore di default è pari alla distanza minima tra coppie di nodi della grafo.

4.5.2 Heat Map

Il grafico Heat Map (figura 10) permette di visualizzare una matrice bidimensionale delle distanze, contenente rettangoli colorati.

Ogni rettangolino posizionato *al di fuori* della diagonale_G rappresenta un arco di un grafo_G. La riga e la colonna di appartenenza indicano ciascuna un nodo che è estremo dell'arco, il cui nome è riportato sopra la riga o la colonna. Ogni rettangolino ha un colore tendente al viola, di opacità proporzionale all'intensità del collegamento tra i nodi rappresentanti riga e colonna.

Ogni rettangolino posizionato *nella* diagonale_G *non* rappresenta un arco di un grafo_G, bensì il nodo rappresentante la riga e la colonna (che hanno lo stesso indice, trattandosi di elementi sulla diagonale). È riempito con un colore opaco diverso dal viola, che rappresenta il gruppo di appartenenza del nodo.

Ciascun collegamento tra nodi è presente due volte nel grafico Heat Map: una volta a destra della diagonale_G, un'altra volta a sinistra della diagonale_G.

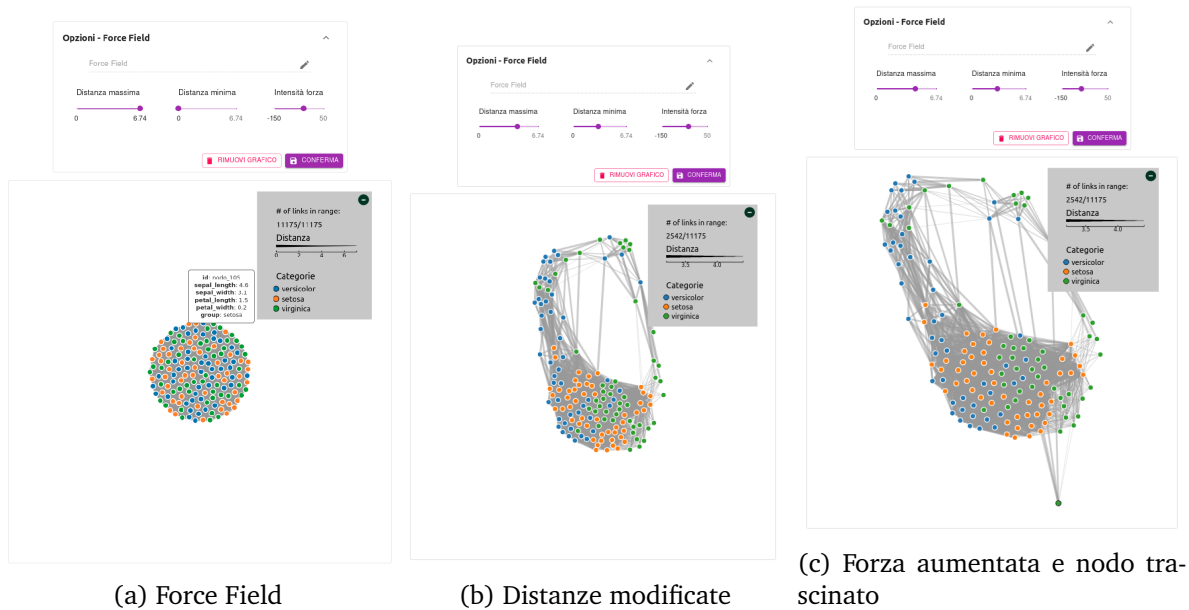


Figura 9: Modifica parametri Force Field

4.5.2.1 Modalità di interazione

Posizionando il cursore sopra un rettangolino è possibile **evidenziare** a lato del grafico il nome dei nodi corrispondenti agli estremi dell'arco rappresentato dal rettangolino.

Posizionando il cursore sopra il nome di una riga o di una colonna è possibile **visualizzare informazioni** sul nodo.

4.5.2.2 Parametri della visualizzazione

Attraverso il pannello di controllo della visualizzazione si possono modificare i seguenti parametri:

- **Ordinamento:** è possibile *ordinare* gli elementi della matrice, selezionando la modalità di ordinamento dall'apposito menu a tendina.
Le opzioni selezionabili sono:
 - **ordinamento per cluster_G**;
 - **ordinamento originale.**
- **Distanza_G massima:** distanza_G massima tra i nodi affinché venga visualizzato nel grafico il rettangolino corrispondente all'arco che collega la coppia di nodi. Il suo valore di default è pari alla distanza massima tra coppie di nodi della grafo.
- **Distanza_G minima:** distanza_G minima tra i nodi affinché venga visualizzato nel grafico il rettangolino corrispondente all'arco che collega la coppia di nodi. Il suo valore di default è pari alla distanza minima tra coppie di nodi della grafo.



Figura 10: Modifica parametri Heat Map

4.5.3 Scatter plot Matrix

Il grafico Scatter plot Matrix (figura 11) permette di visualizzare una griglia di Scatter plot_G, ciascuno avente come assi una coppia di dimensioni del dataset. Ciò significa che, preso un dataset con k dimensioni, lo Scatter plot Matrix avrà k righe e k colonne, e il grafico alla i -esima riga e alla j -esima colonna è uno Scatter plot_G con la i -esima dimensione nell'asse delle ascisse e la j -esima dimensione nell'asse delle ordinate.

È possibile visualizzare fino ad un massimo di 5 dimensioni nello Scatter plot Matrix.

N.B. Ogni coppia di dimensioni distinte appare in due diversi Scatter plot_G della matrice, in posizioni speculari rispetto alla diagonale, con assi delle ascisse e delle ordinate invertiti.

4.5.3.1 Modalità di interazione

L'operazione di *brushing* permette di selezionare un insieme di punti in uno Scatter plot_G e di vedere evidenziati quei punti in tutti i restanti Scatter plot_G della matrice.

È possibile *effettuare il brush_G* dei nodi **selezionando un area rettangolare** in uno Scatter plot_G della griglia. In seguito a questa operazione in ogni Scatter plot_G sono distinguibili:

- i punti selezionati, nelle dimensioni e nei colori originari;
- i punti non selezionati, in dimensioni nettamente ridotte e in colore nero.

È possibile *spostare* il rettangolo di selezione precedentemente disegnato, **trascinandolo** con il mouse nella posizione desiderata. L'evidenziamento dei dati nei restanti Scatter plot viene aggiornato in tempo reale.

È possibile *annullare il brush_G* della selezione precedentemente creata, **clickando in un punto** del grafico **al di fuori** del rettangolo di selezione. In seguito a questa operazione tutti i punti vengono visualizzati nella loro forma originaria e non sono più visibili aree di selezione nel grafico.

È possibile *aggiungere o rimuovere* una dimensione selezionandola dal pannello in parte al grafico. Ad ogni aggiunta o rimozione di una dimensione l'eventuale brush_G applicato viene rimosso.

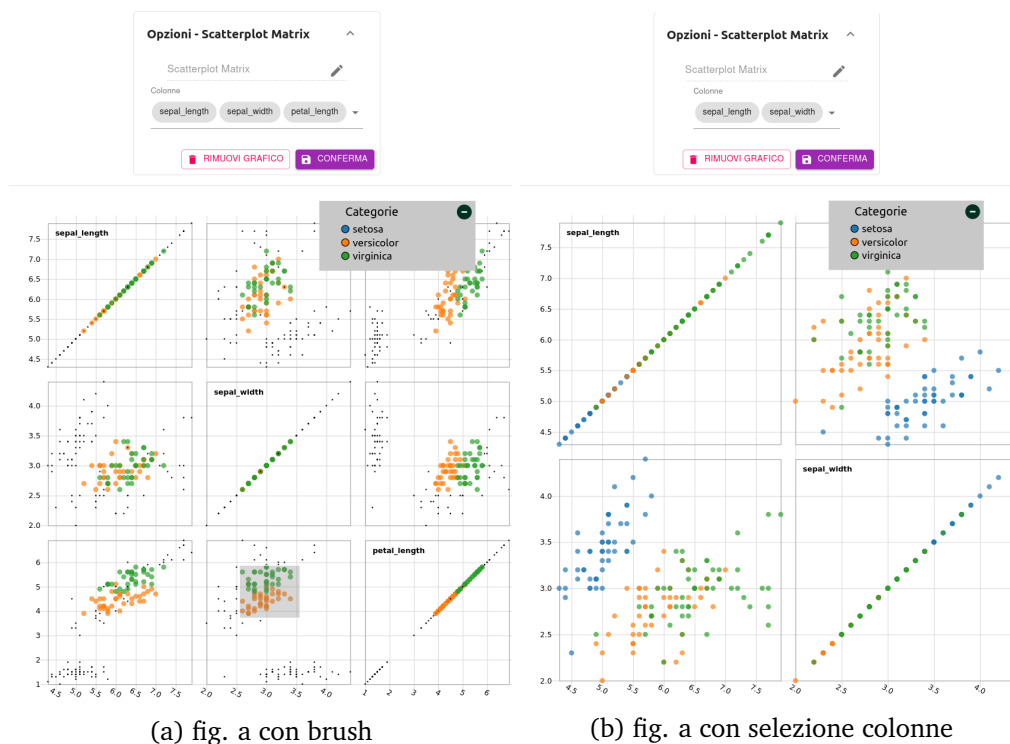


Figura 11: Interazione con Scatter plot Matrix

4.5.4 Proiezione Lineare Multi Asse

La Proiezione Lineare Multi Asse (figura 12) è una variante dello Scatter plot_G tradizionale, nella quale gli assi sono visibili e trascinabili.

Il trascinamento di un asse mostra a schermo una differente proiezione_G bi-dimensionale del dataset multi-dimensionale_G originario.

L'asse delle ascisse e delle ordinate rappresentano le componenti principali 1 e 2. Una griglia posiziona in secondo piano rispetto al grafico aiuta a dedurre le coordinate dei punti rispetto alle due componenti principali. L'origine_G è rappresentata dall'intersezione degli assi.

4.5.4.1 Modalità di interazione

È possibile *visualizzare una diversa proiezione_G* del dataset originario **trascinando** l'estremità di un asse in un altro punto del grafico avente distanza_G uguale (o simile) dall'origine_G.

È possibile *comprimere o allungare* un asse **trascinando** l'estremità dell'asse in un altro punto del grafico più vicino o lontano dall'origine_G, senza effettuare rotazioni.

Un'operazione di trascinamento può produrre uno solo o entrambi gli effetti descritti sopra, poiché l'utente non è limitato a muovere un'estremità di un asse lungo una stessa orbita circolare o lungo la stessa direzione più o meno lontana dall'origine_G.

È possibile *aggiungere o rimuovere* un asse selezionandolo dal pannello in parte al grafico. Ad ogni aggiunta o rimozione di un asse la proiezione_G viene ricalcolata, dunque gli assi già presenti o rimasti potrebbero assumere una posizione differente.

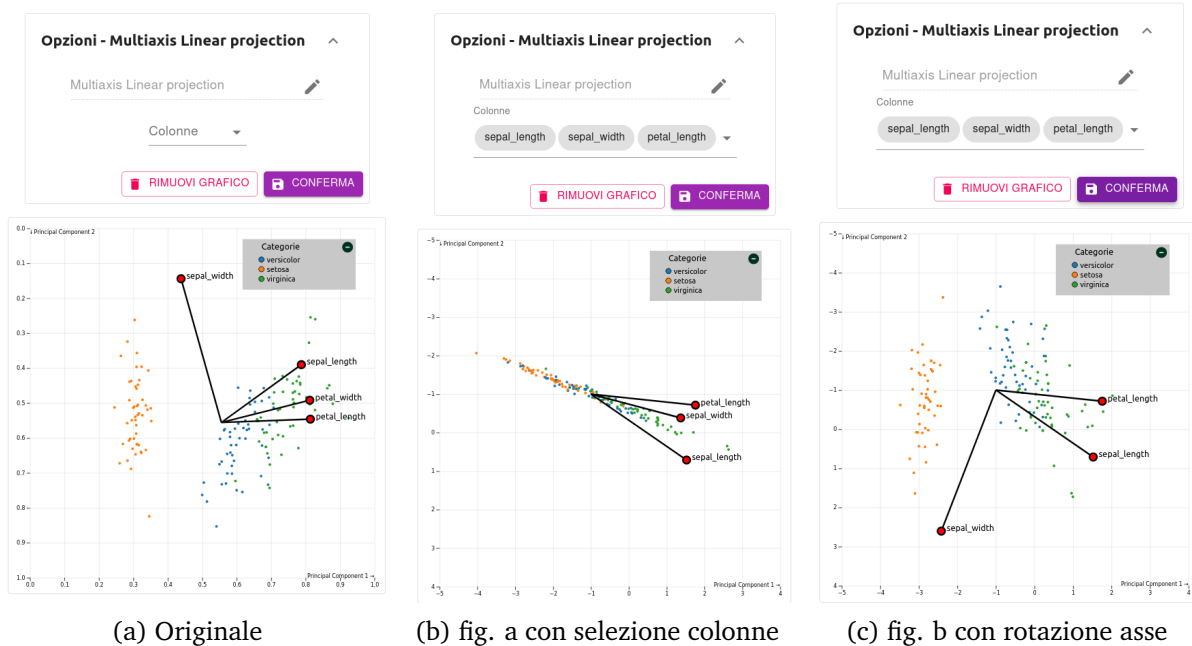


Figura 12: Interazione con Proiezione Lineare Multi Asse

4.6 Complementi ai grafici

4.6.1 Legenda

Al fine di semplificare l'attività esplorativa sui dati ogni grafico possiede una propria legenda, visibile di default e nascondibile.

Per ogni tipologia di grafico la legenda mostra le *categorie* associate ai dati visualizzati nel grafico. Nei grafici *Force Field* e *Heat Map* vengono mostrate inoltre la *scala della distanza* adottata, il *numero di archi* del grafo compresi tra il *valore della distanza minima e massima* e, se sono presenti archi con valore della distanza nullo, il *numero di archi* del grafo aventi valore della *distanza nullo* (una funzione di calcolo delle distanze può produrre valori nulli, si veda §4.3 - Scelta di una funzione per il calcolo delle distanze).

Nei grafici *Scatter plot Matrix* e *Proiezione Lineare Multi Asse* viene visualizzato il numero di elementi del dataset non mostrati poiché contenenti valori nulli (una funzione di calcolo delle distanze può produrre valori nulli, si veda §4.2 - Scelta di una funzione di riduzione dimensionale).

A Glossario

Di seguito non sono riportate le definizioni di funzioni matematiche e algoritmi per i quali si è ritenuto poco utile fornire una definizione concisa e povera di formalismi matematici che si adattasse a un glossario.

Si rimanda dunque a §1.4 - Riferimenti per la definizione dei seguenti termini:

- ALGORITMI DI RIDUZIONE DIMENSIONALE_G: FASTMAP, ISOMAP, LLE, t-SNE, UMAP;
- FUNZIONI DI CALCOLO DELLE DISTANZE_G: distanza di Canberra, distanza di Chebyshev, distanza Cosine, distanza Euclidea, distanza Euclidea Quadrata, distanza di Manhattan.

B

Browser

Applicazione per la fruizione di contenuti web.

Brush

In uno Scatter plot Matrix, l'operazione di Brush permette di selezionare un insieme di punti in uno Scatter plot e di vedere evidenziati quei punti in tutti i restanti Scatter plot della matrice.

Bug

Guasto che porta al malfunzionamento del software, tipicamente dovuto ad un errore nella scrittura del codice sorgente di un programma.

Build

Trasformazione del codice sorgente in codice eseguibile.

C

Checkbox

Controllo grafico con cui l'utente può effettuare selezioni multiple.

Cluster

Gruppo di oggetti di un insieme. Può essere dato in partenza o può essere ricavato da particolari algoritmi di apprendimento non supervisionato, detti di *clustering*.

CSV

Acronimo di *Comma-Separated Values*, è un formato di file utilizzato per la rappresentazione di una tabella di dati, che utilizza il carattere delimitatore virgola ',' o punto e virgola ';' per separare le celle.

D

Database

Insieme organizzato di dati, gestito da un DBMS (DataBase Management System).

Dati multi-dimensionali

Dataset in cui il numero di features per ogni record del dataset è $>> 3$.

Diagonale

In una matrice A , la diagonale è l'insieme di elementi $A_{i,j}$ tali che $i = j$.

Distanza

Misura della "lontananza" tra due punti di un insieme al quale si possa attribuire qualche carattere spaziale.

DSV

Acronimo di *Delimiter-Separated Values*, è una famiglia di formati di file utilizzati per la rappresentazione di una tabella di dati, che utilizzano un carattere delimitatore per separare le celle. CSV_G e TSV_G sono formati DSV.

E**EDA**

Acronimo di *Exploratory Data Analysis*, parte dell'attività di Data Analysis, nella quale l'analista ricerca associazioni, sequenze ripetute nascoste o pattern nei dati.

F**Forza**

Grandezza fisica vettoriale che si manifesta nell'interazione reciproca di due o più corpi sia a livello macroscopico, sia a livello delle particelle elementari.

G**Grafo**

In Matematica Discreta, un grafo è una tripla (V, E, f) , dove V è detto insieme dei nodi, E è detto insieme degli archi e f è una funzione che associa ad ogni arco e in E due vertici u, v in V .

Git

Sistema di versionamento $_G$ distribuito multi-piattaforma. Permette di versionare i sorgenti software $_G$, documenti di testo e di collaborare alla loro realizzazione.

GitHub

Servizio di hosting per progetti software $_G$, che implementa lo strumento di controllo versione distribuito Git_G .

I

Integratore di Verlet

Metodo numerico per il calcolo delle Equazioni del moto di Newton.

M

Matrice delle distanze

Matrice quadrata contenente le distanze tra ogni coppia di un insieme.

N

Node.js

Framework open-source multi-piattaforma orientato agli eventi per l'esecuzione di codice JavaScript.

npm

Acronimo di *Node.js Package Manager*, gestore di pacchetti predefinito per Node.js.

O

Origine

Nel sistema di riferimento cartesiano, il punto nel quale si intersecano tutte le rette orientate dette *assi*.

P

PostgreSQL

Database ad oggetti e relazionale open-source, focalizzato sulla scalabilità e sulla massima aderenza allo standard SQL.

Proiezione

Trasformazione lineare definita da uno spazio vettoriale in sé stesso.

R

Radio button

Controllo grafico che consente all'utente di effettuare una scelta singola esclusiva nell'ambito di un insieme predefinito di opzioni o possibili scelte.

Repository

Archivio o sito web nel quale sono raccolti e conservati dati ed informazioni in formato digitale.

Riduzione dimensionale

Algoritmo di trasformazione di dati da uno spazio multi-dimensionale_G ad uno spazio con meno dimensioni. Un algoritmo di riduzione dimensionale cerca di mantenere le caratteristiche dello spazio multi-dimensionale_G anche nello spazio con meno dimensioni.

S

Scatter plot

Tipo di grafico in cui due variabili di un set di dati sono riportate su uno spazio cartesiano. Viene chiamato anche grafico di dispersione.

Scatter plot Matrix e Proiezione Lineare multi asse sono varianti di Scatter plot.

Server

Componente o sottosistema informatico di elaborazione e gestione del traffico di informazioni che fornisce, a livello logico e fisico, un qualunque tipo di servizio ad altre componenti che ne fanno richiesta attraverso una rete di computer, all'interno di un sistema informatico o anche direttamente in locale su un computer.

Slider

Componente grafico con il quale un utente può impostare un valore muovendo un indicatore, solitamente con uno spostamento orizzontale.

SVG

Acronimo di *Scalable Vector Graphics*, indica un particolare formato che è in grado di visualizzare oggetti di grafica vettoriale e quindi di salvare immagini in modo che siano ridimensionate a piacere senza perdere in risoluzione grafica.

T

TSV

Acronimo di *Tab-Separated Values*, è un formato di file utilizzato per la rappresentazione di una tabella di dati, che utilizza il carattere delimitatore tabulazione per separare le celle.

V

Validatore

Un validatore è uno strumento in grado di controllare la conformità di un file rispetto a uno standard.