

CODEBUSTERS

Progetto: *HD Viz*
codebusterswe@gmail.com

Manuale Utente

Informazioni sul documento

Versione	1.0.0-1.7
Approvatori	Rago Alessandro Sassaro Giacomo
Redattori	Scialpi Paolo Rago Alessandro Safdari Hossain
Verificatori	Pirolo Alessandro Zenere Marco Baldiiseri Michele
Uso	Interno <i>Zucchetti</i>
Distribuzione	Prof. Vardanega Tullio Prof. Cardin Riccardo Gruppo <i>CodeBusters</i>

Descrizione

Questo documento racchiude le istruzioni per l'utilizzo corretto e consapevole del prodotto *HD Viz*.

Registro delle modifiche

Versione	Data	Nominativo	Ruolo	Descrizione
1.0.0-1.7	22-04-2021	Rago Alessandro	Responsabile	Approvazione del documento
0.2.0-1.7	19-04-2021	Baldisseri Michele	Verificatore	Verifica complessiva del documento
0.1.3-1.7	18-04-2021	Safdari Hossain, Baldisseri Michele	Programmatore, Verificatore	Aggiornata §4, stesura §5 e verifica
0.1.2-1.7	17-04-2021	Safdari Hossain, Sassaro Giacomo	Programmatore, Verificatore	Aggiornate immagini, aggiornata §2 e verifica
0.1.1-0.6	27-03-2021	Sassaro Giacomo, Safdari Hossain	Programmatore, Verificatore	Stesura §A e verifica
0.1.0-0.5	24-03-2021	Scialpi Paolo	Verificatore	Verifica complessiva del documento
0.0.3-0.5	21-03-2021	Baldisseri Michele, Safdari Hossain	Programmatore, Verificatore	Modifica sezione §3, inserite immagini in §4, stesura §6 e verifica
0.0.2-0.4	20-03-2021	Sassaro Giacomo, Pirolo Alessandro	Amministratore, Verificatore	Stesura delle sezioni §1, §2, §3, §4 e verifica
0.0.1-0.4	19-03-2021	Sassaro Giacomo, Zenere Marco	Amministratore, Verificatore	Creazione scheletro documento e verifica

Indice

1	Introduzione	5
1.1	Scopo del documento	5
1.2	Scopo del prodotto	5
1.3	Glossario	5
2	Requisiti minimi di sistema	6
2.1	Requisiti di sistema	6
2.2	Requisiti hardware	6
2.3	Browser	7
3	Installazione	8
3.1	Clonare la repository	8
3.2	Importare il database	8
3.3	Avviare il server	8
3.4	Avviare la web app	9
4	Istruzioni all'uso	10
4.1	Caricamento dati	10
4.2	Riduzione dimensionale	14
4.3	Calcolo della distanza	15
4.4	Scelta del grafico e visualizzazione dei dati	16
4.4.1	Scatterplot Matrix	17
4.4.2	Adjacency Matrix	19
4.4.3	Heat Map	21
4.4.4	Force Field	22
4.4.5	Proiezione Lineare Multi Asse	23
4.5	Formattazione dei file	24
5	Video Tutorial	25
5.1	Caricamenti dati da file CSV	25
5.2	Caricamenti dati da Database	25
5.3	Riduzione dimensionale	25
5.4	Calcolo delle distanze	25
5.5	Scatterplot Matrix	25
5.6	Adjacency Matrix	25
5.7	Heat Map	25
5.8	Force Field	25
5.9	Proiezione lineare Multi Asse	25
6	Supporto tecnico	26
A	Glossario	27

Elenco delle tabelle

1	Requisiti di sistema	6
2	Requisiti hardware	6
3	Prestazioni ottimali	6
4	Browser e versioni compatibili	7

Elenco delle figure

1	Avvio dell'applicazione	9
2	Scelta del tipo di caricamento	10
3	Caricamento dati da database	11
4	Caricamento dati da file CSV	12
5	Messaggi d'esito del caricamento dati nel sistema	12
6	Riduzione dimensionale o calcolo della distanza	13
7	Finestra per la riduzione dimensionale tramite algoritmo	14
8	Finestra per la riduzione dimensionale tramite il calcolo delle distanze	15
9	Lista dei grafici che è possibile scegliere	16
10	Form delle preferenze per il grafico Scatterplot Matrix	17
11	Esempio di visualizzazione dei dati con il grafico Scatterplot Matrix	18
12	Form delle preferenze per il grafico Adjacency Matrix	19
13	Esempio di visualizzazione dei dati con il grafico Adjacency Matrix	20
14	Form delle preferenze per il grafico Heat Map	21
15	Esempio di visualizzazione dei dati con il grafico Heat Map	21
16	Form delle preferenze per il grafico Force Field ed esempio di visualizzazione	22
17	Form delle preferenze per il grafico PLMA ed esempio di visualizzazione	23
18	Esempio di file CSV correttamente formattato	24

1 Introduzione

1.1 Scopo del documento

Questo documento ha lo scopo di illustrare le istruzioni per l'utilizzo e le funzionalità fornite dall'applicazione. L'utente sarà quindi a conoscenza dei requisiti minimi necessari per il corretto funzionamento di *HD Viz*, di come installarla in locale e di come farne un utilizzo consapevole.

1.2 Scopo del prodotto

Oggigiorno anche i programmi più tradizionali gestiscono e memorizzano una grande mole di dati; di conseguenza servono software in grado di eseguire un'analisi e un'interpretazione delle informazioni. Il prodotto HDViz ha come obiettivo quello di creare un'applicazione di visualizzazione di dati con numerose dimensioni in modo da renderle comprensibili all'occhio umano. Lo scopo del prodotto sarà quello di fornire all'utente diversi tipi di visualizzazioni e di algoritmi per la riduzione dimensionale in modo che, attraverso un processo esplorativo, l'utilizzatore del prodotto possa studiare tali dati ed evidenziarne degli eventuali cluster^G.

1.3 Glossario

Per evitare ambiguità relative alle terminologie specifiche al dominio di utilizzo, queste verranno evidenziate da una 'G' ad apice e riportate nel glossario presente nell'appendice §A.

2 Requisiti minimi di sistema

Per poter utilizzare l'applicazione web HDViz è necessario soddisfare i seguenti requisiti minimi.

2.1 Requisiti di sistema

Per far sì che le operazioni di installazione e avvio del prodotto avvengano correttamente e che si possa aver accesso a tutte le funzionalità, è necessario avere nella propria macchina i seguenti software.

Software	Versione	Riferimento per il download
Node.js	14.16.x	https://nodejs.org/it/
Npm	7.x	Integrato nel download di Node.js
PostgreSQL	13.x	https://www.postgresql.org/download/

Tabella 1: Requisiti di sistema

2.2 Requisiti hardware

Per avere delle prestazioni accettabili dell'applicazione è preferibile avere almeno i seguenti componenti hardware.

Componente	Requisito
Processore	Quad-Core 3,2 GHz
RAM	8GB DDR4

Tabella 2: Requisiti hardware

La connessione internet ideale dovrebbe essere di almeno 80Mb/s in download per assicurare i seguenti tempi di risposta (in secondi) della web app.

Tempo	Operazione
2	Caricamento di un dataset di 2Mb;
7	Applicazione di un algoritmo di riduzione dimensionale con 4 dimensioni e 500 punti in totale;
2	Visualizzazione di un grafico con 500 punti.

Tabella 3: Prestazioni ottimali

2.3 Browser

L'applicazione è stata testata e quindi resa compatibile con le ultime versioni dei browser maggiormente utilizzati al momento.

Browser	Versione
Chrome	87
Edge	79
Mozilla Firefox	84
Safari	13.1

Tabella 4: Browser e versioni compatibili

3 Installazione

Per utilizzare l'applicazione web è necessario:

- Clonare la repository (**Obbligatorio**);
- Importare il database (*Opzionale*);
- Avviare il server (*Opzionale*);
- Avviare la web app (**Obbligatorio**).

3.1 Clonare la repository

- 1) Scaricare il codice come file .zip direttamente dal repository CodeBusters-HDViz:

<https://github.com/CodeBusterswe/CodeBusters-HDViz>

Oppure, con Git^G installato in locale, è possibile clonare il repository con il comando:

git clone https://github.com/CodeBusterswe/CodeBusters-HDViz

- 2) Localizzare da terminale la cartella in cui si è stato estratto/clonato il prodotto:

cd percorso\CodeBusters-HDViz

3.2 Importare il database

- 1) Installare PostgreSQL e pgAdmin dal sito:

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

- 2) Creare un nuovo database;
- 3) Fare il restore del database utilizzando il file **dumbDB.js** presente nella cartella `server` del prodotto.

3.3 Avviare il server

- 1) Entrare nella cartella `server` con il comando:

cd server

- 2) Nel caso di primo avvio, digitare:

npm install

- 3) Controllare che i valori impostati nel file **default.js** nella sottocartella `config` corrispondano a quelli impostati in PostgreSQL:

- 'USERNAME' : nome utente dell'account;
- 'PASSWORD' : password dell'account;
- 'HOST' : porta di postgresSQL;
- 'DB_NAME' : nome del database precedentemente creato;

4) Avviare il server con il comando:

npm start

3.4 Avviare la web app

1) Entrare nella cartella `client` con il comando:

cd client

2) Nel caso di primo avvio digitare il comando:

npm install

3) Fare la build dell'app con il comando:

npm run build

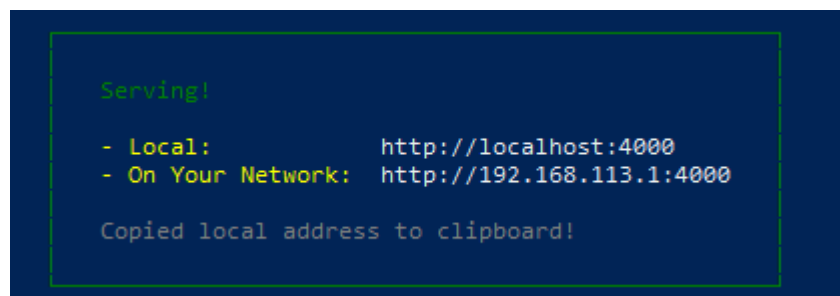
4) nel caso di primo avvio dell'applicazione sul sistema, digitare:

npm install -g serve

5) Digitare poi:

serve -s build -l 4000

L'applicazione sarà disponibile aprendo l'indirizzo fornito dal terminale.



```
Serving!
- Local:      http://localhost:4000
- On Your Network: http://192.168.113.1:4000

Copied local address to clipboard!
```

Figura 1: Avvio dell'applicazione

4 Istruzioni all'uso

Tutte le funzionalità di *HD Viz* sono facilmente reperibili nel menu laterale a comparsa. Per aprirlo basta semplicemente passarci sopra con il cursore.

4.1 Caricamento dati

Inizialmente sono utilizzabili solo due voci. Le restanti infatti sono disabilite perché necessitano che si siano prima caricati dei dati, indifferentemente dalla fonte. Ciò è comunque segnalato a video con un tooltip passando sopra a tali voci con il cursore. Quelle disponibili sono:

- Carica dati dal database^G;
- Carica dati da CSV^G;

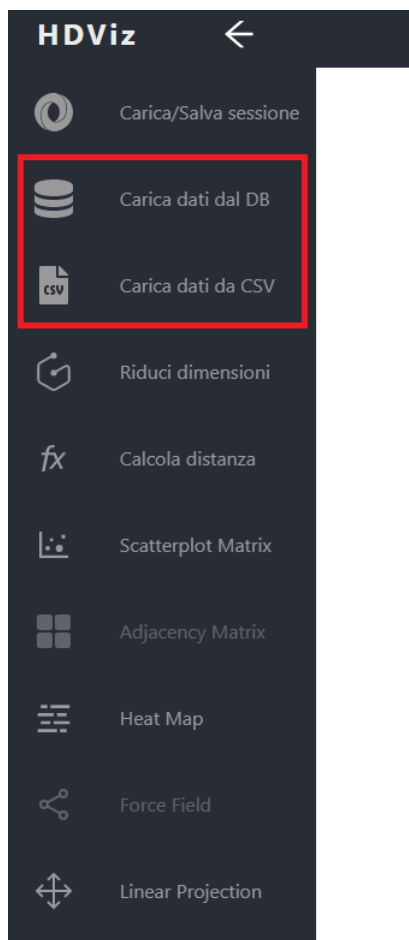


Figura 2: Scelta del tipo di caricamento

Entrambe aprono una finestra di dialogo per compiere le operazioni indicate. In particolare:

- La finestra per il caricamento dei dati dal database^G permette di scegliere uno dei dataset presenti ed effettuare immediatamente una selezione delle dimensioni da caricare.

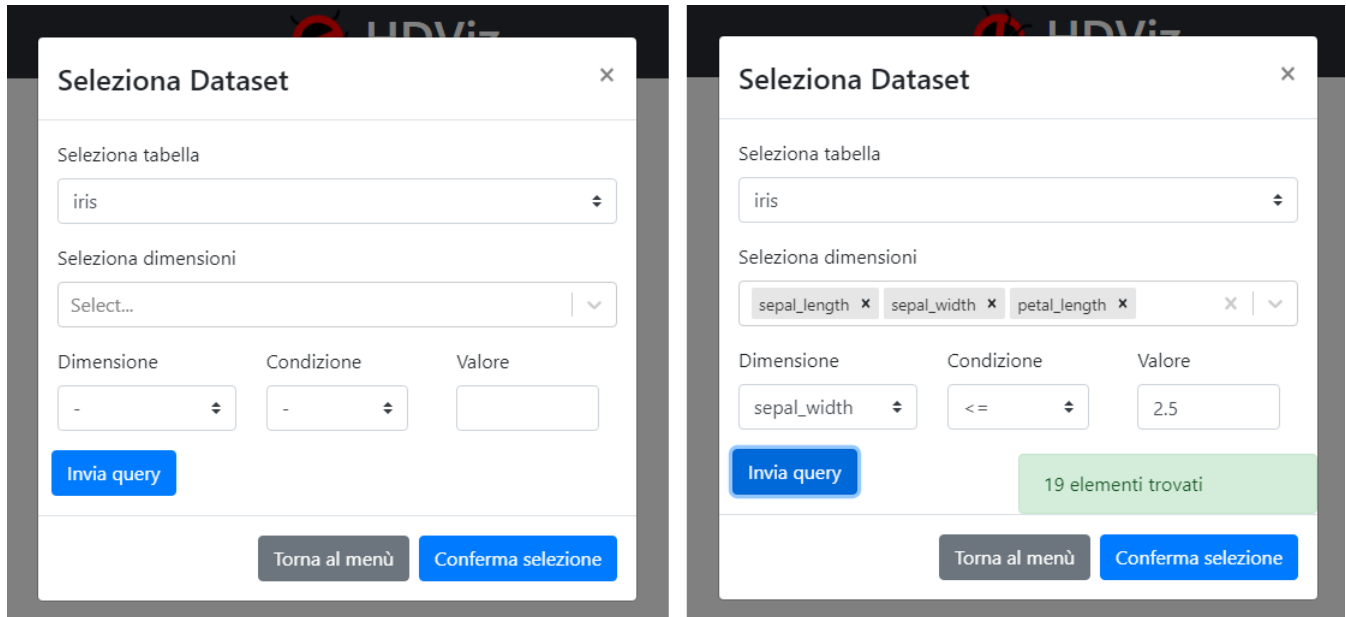


Figura 3: Caricamento dati da database

Dopo aver selezionato le dimensioni desiderate dal dataset scelto, premendo il tasto "Invia query" verrà visualizzato un messaggio d'esito (positivo o negativo) contenente il numero di elementi individuati. L'utente ha la possibilità d'imporre delle condizioni su determinate dimensioni, in modo da prelevare solo particolari punti di suo interesse.

Terminata la scrematura dei dati, premendo il tasto "Conferma selezione" i dati verranno caricati nel sistema.

- La finestra per il caricamento dei dati da file CSV^G presenta una sezione *drag and drop*^G per il caricamento di un file .csv dell'utente; in alternativa, la sezione può essere cliccata per utilizzare l'esplora risorse del proprio dispositivo. Caricato il file, la finestra permette la selezione delle dimensioni che si desidera utilizzare. Di default tutte le dimensioni presenti nel dataset vengono preselezionate e premendo il tasto "Conferma selezione" i dati saranno caricati nel sistema.

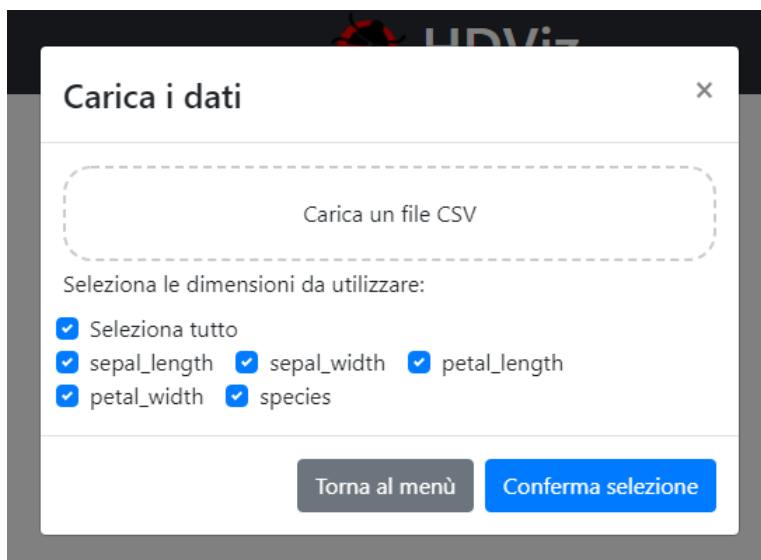


Figura 4: Caricamento dati da file CSV

Di seguito sono riportati i messaggi d'esito, positivo o negativo, in riferimento all'effettivo caricamento dei dati nel sistema e che saranno mostrati all'utente alla chiusura dei modali spiegati precedentemente.

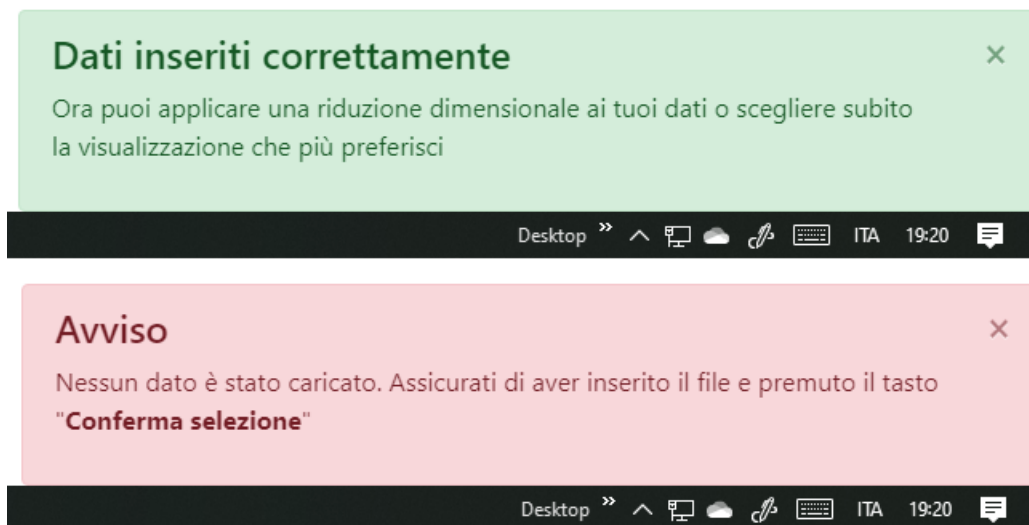


Figura 5: Messaggi d'esito del caricamento dati nel sistema

Una volta caricati i dati è possibile procedere con la loro elaborazione. A questo scopo sono disponibili le seguenti voci:

- Riduci dimensioni;
- Calcola distanza.

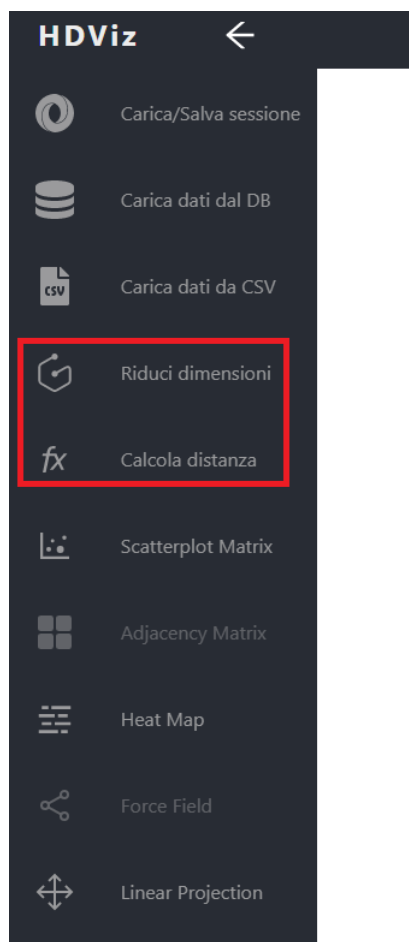


Figura 6: Riduzione dimensionale o calcolo della distanza

In ogni caso l'applicazione di algoritmi di riduzione dimensionale o funzioni per il calcolo della distanza **non sono obbligatorie** per la maggior parte delle visualizzazioni dei dati. Solo Adjacency Matrix e Force Field richiedono necessariamente che sia stata prima caricata una matrice delle distanze (attraverso il calcolo della distanza). Se non si è interessati ai suddetti algoritmi si può passare direttamente a scegliere il grafico che più si preferisce tra quelli proposti.

4.2 Riduzione dimensionale

Cliccando sulla voce "Riduci dimensioni", si apre una finestra che permette di scegliere quali dimensioni saranno interessate dal processo e quale algoritmo utilizzare tra:

- FastMap^G;
- LLE^G;
- IsoMap^G;
- t-SNE^G.

In base a questa ultima scelta l'utente potrà selezionare una serie di parametri specifici per eseguire la riduzione come più preferisce.

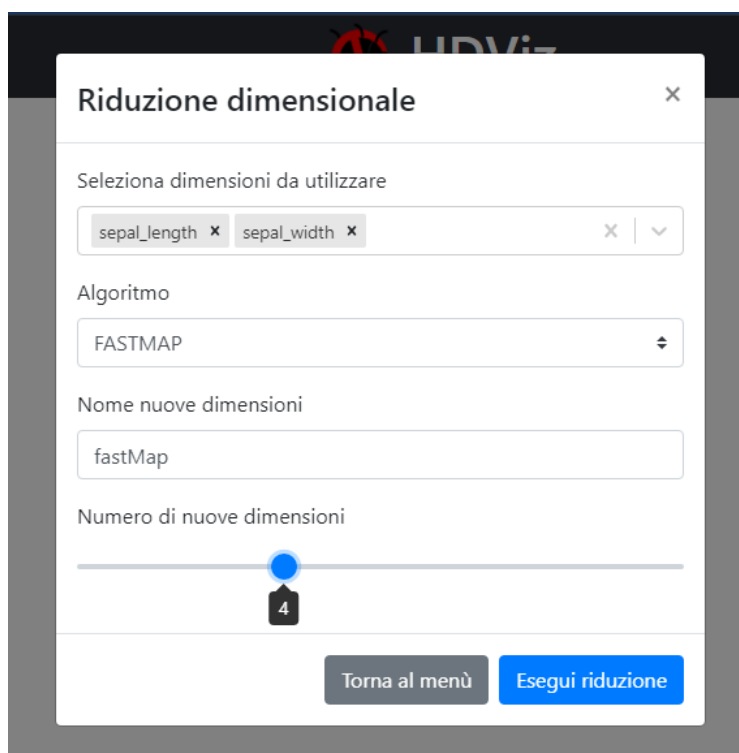


Figura 7: Finestra per la riduzione dimensionale tramite algoritmo

Ultimata la configurazione, premendo il tasto "Esegui riduzione" verranno create ed inserite le nuove dimensioni nel sistema.

4.3 Calcolo della distanza

Cliccando sulla voce "Calcola distanza", si apre una finestra che permette di scegliere quali dimensioni utilizzare per creare una matrice delle distanze, quale nome assegnarvi e quale funzione di distanza usare tra:

- Euclidean^G;
- Canberra^G;
- Chebyshev^G;
- Manhattan^G.

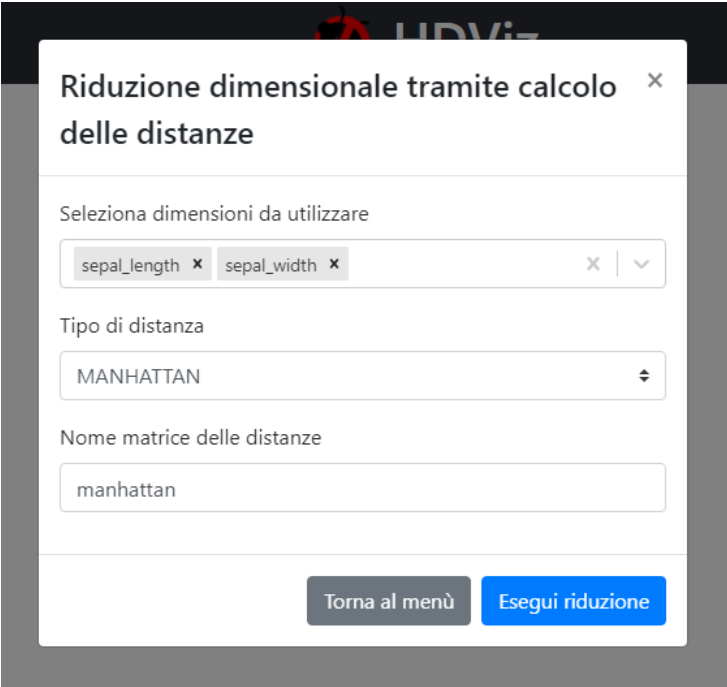
The image shows a web application window titled "Riduzione dimensionale tramite calcolo delle distanze" with a close button (X) in the top right corner. The window contains three main sections: 1. "Seleziona dimensioni da utilizzare" with a multi-select box containing "sepal_length" and "sepal_width", each with a small 'x' to remove it, and a dropdown arrow on the right. 2. "Tipo di distanza" with a dropdown menu currently showing "MANHATTAN" and a double-headed arrow icon. 3. "Nome matrice delle distanze" with a text input field containing the word "manhattan". At the bottom right, there are two buttons: "Torna al menù" (grey) and "Esegui riduzione" (blue).

Figura 8: Finestra per la riduzione dimensionale tramite il calcolo delle distanze

Terminata la configurazione, premendo il tasto "Esegui riduzione" verrà creata la matrice e salvata nel sistema. Le matrici create dall'utente saranno disponibili in tutti i grafici che dipendono dal concetto di distanza (come Adjacency Matrix e Force Field).

4.4 Scelta del grafico e visualizzazione dei dati

Una volta elaborati i dati o eventualmente immediatamente dopo al caricamento dei dati, si può scegliere il grafico che più si preferisce tra quelli proposti:

- Scatterplot Matrix^G;
- Adjacency Matrix^G;
- Heat Map^G;
- Force Field^G;
- PLMA (Proiezione lineare multi asse)^G.

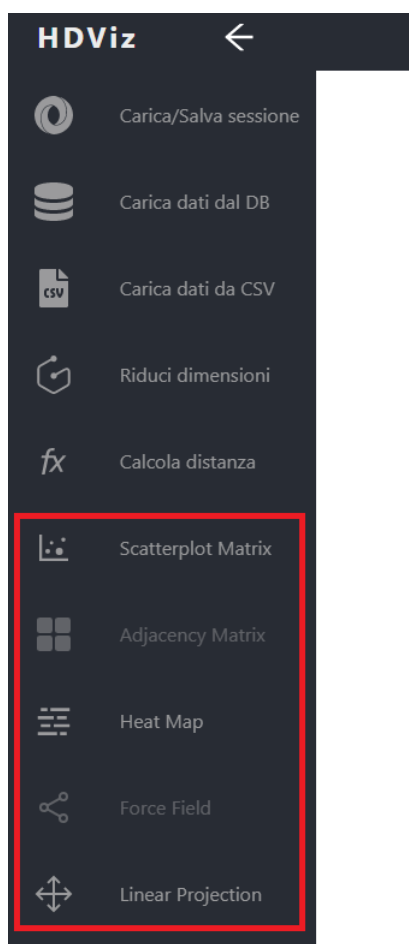


Figura 9: Lista dei grafici che è possibile scegliere

Una volta selezionata una di queste opzioni, si aprirà una form^G sulla destra attraverso la quale sarà possibile modificare la visualizzazione del grafico. Inizialmente tutti i campi sono settati a "Nessuna dimensione" quindi è corretto che il grafico non sia già visualizzabile. È necessario modificare tali campi a proprio piacimento. Ad ogni loro modifica la visualizzazione cambierà dinamicamente, adattandosi alle dimensioni del dispositivo. La form è inoltre sempre accompagnata da un bottone per nasconderla e centralizzare il grafico nello schermo per concentrarsi solamente sull'analisi del grafico.

4.4.1 Scatterplot Matrix

Per questa visualizzazione la form^G a destra permette di:

- Modificare le dimensioni da applicare agli assi;
- Modificare la dimensione per l'applicazione del colore sui punti.

Nascondi preferenze

Axis one

sepal_length

Axis two

sepal_width

Axis Three

petal_length

Axis Four

petal_width

Axis Five

species

Color

species

Figura 10: Form delle preferenze per il grafico Scatterplot Matrix

Di seguito un esempio di visualizzazione dei dati con questo grafico.

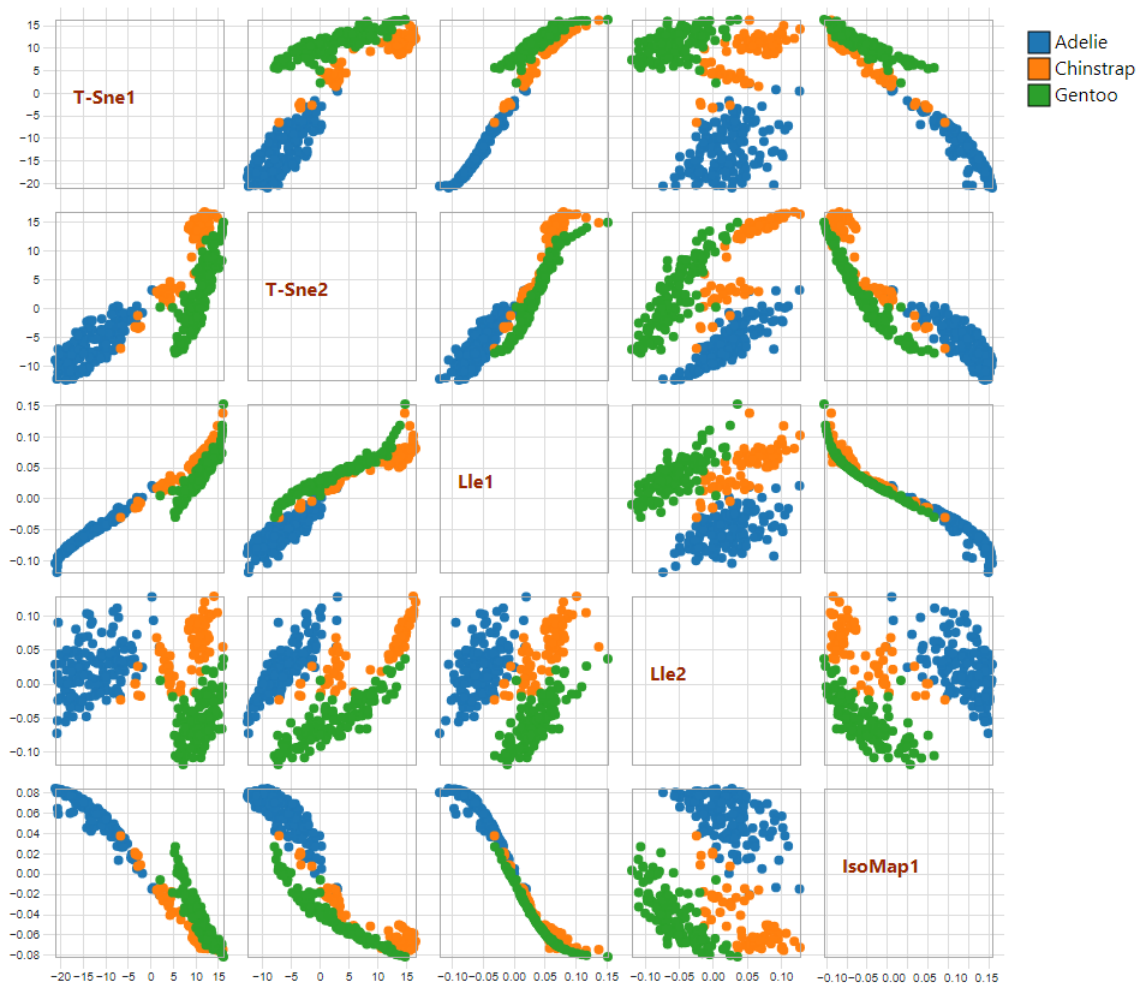


Figura 11: Esempio di visualizzazione dei dati con il grafico Scatterplot Matrix

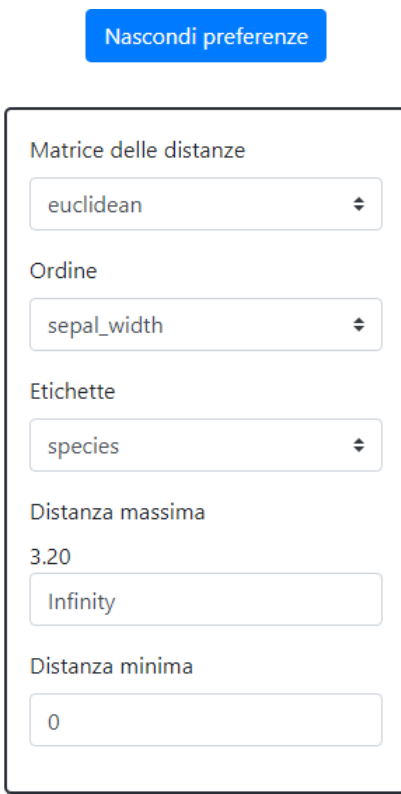
Come visibile in figura, in alto a destra sarà sempre disponibile una legenda dei colori per aiutare l'utente nell'analisi del grafico.

NB: la legenda sarà disponibile solo dopo aver associato una dimensione al colore dei punti dalla form delle preferenze.

4.4.2 Adjacency Matrix

Per questa visualizzazione la form^G a destra permette di:

- Se presente, scegliere una delle matrici delle distanze create dall'utente;
- Scegliere per quale dimensione ordinare i punti del grafico;
- Scegliere quale dimensione associare alle etichette del grafico;
- La distanza minima e massima tra i punti.



The image shows a user preference form for the Adjacency Matrix visualization. At the top is a blue button labeled "Nascondi preferenze". Below it is a white form box with a dark border. Inside the box, there are several sections: "Matrice delle distanze" with a dropdown menu showing "euclidean"; "Ordine" with a dropdown menu showing "sepal_width"; "Etichette" with a dropdown menu showing "species"; "Distanza massima" with a text input field showing "3.20" and a button labeled "Infinity"; and "Distanza minima" with a text input field showing "0".

Figura 12: Form delle preferenze per il grafico Adjacency Matrix

Di seguito un esempio di visualizzazione dei dati con questo grafico.

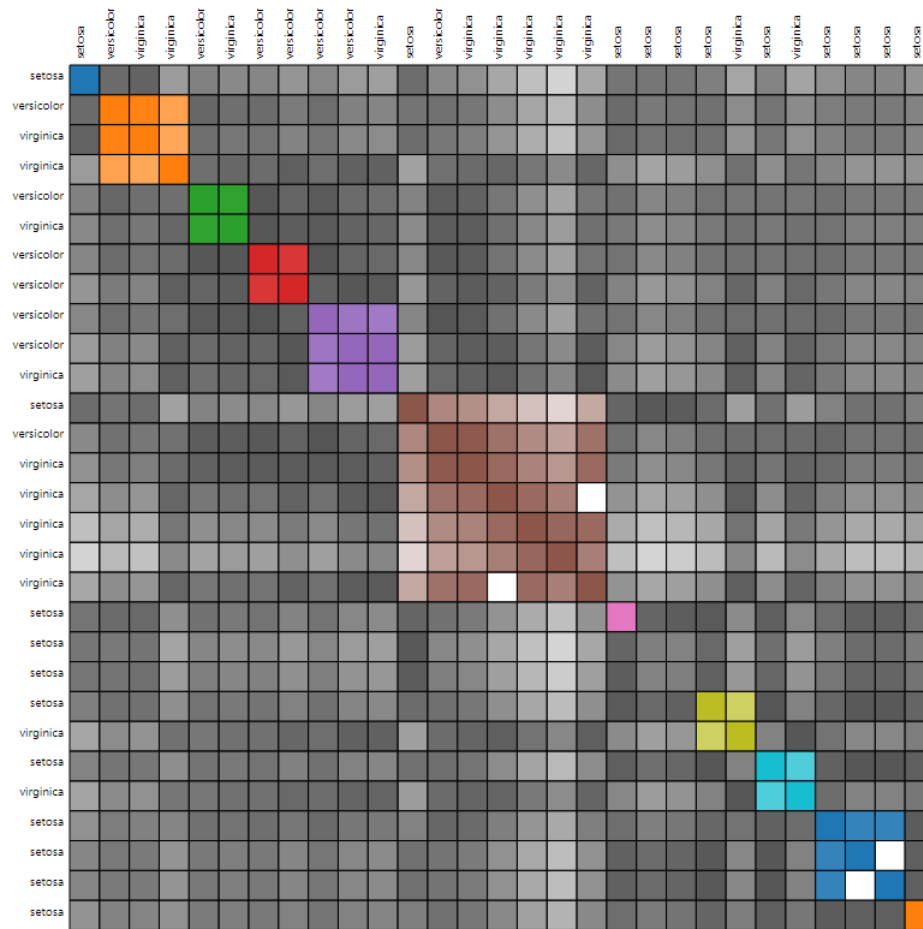


Figura 13: Esempio di visualizzazione dei dati con il grafico Adjacency Matrix

4.4.3 Heat Map

Per questa visualizzazione la form^G a destra permette di:

- Modificare le dimensioni da applicare agli assi X e Y ;
- Modificare la dimensione per l'applicazione della scala di colori.

Nascondi preferenze

Asse X

species

Asse Y

sepal_width

Colore

sepal_width

Figura 14: Form delle preferenze per il grafico Heat Map

Di seguito un esempio di visualizzazione dei dati con questo grafico.

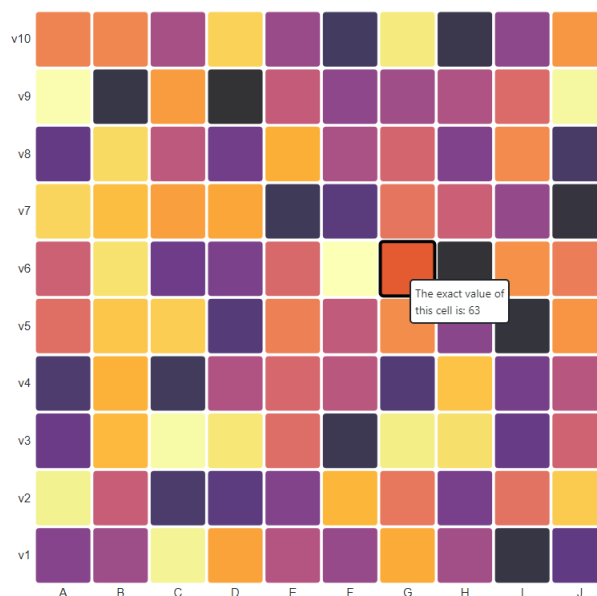


Figura 15: Esempio di visualizzazione dei dati con il grafico Heat Map

Come visibile nella figura precedente, l'utente potrà visualizzare il valore di ogni cella posizionandosi sopra con il cursore. Tale finestra seguirà il movimento del cursore stesso, per sparire una volta posizionato al di fuori del grafico.

4.4.4 Force Field

Per questa visualizzazione la form^G a destra permette di:

- Se presente, scegliere una delle matrici delle distanze create dall'utente;
- Scegliere quale dimensione associare al colore dei nodi del grafico;
- La distanza minima e massima tra i punti.

Di seguito un esempio di visualizzazione dei dati con questo grafico e relativa form delle preferenze.

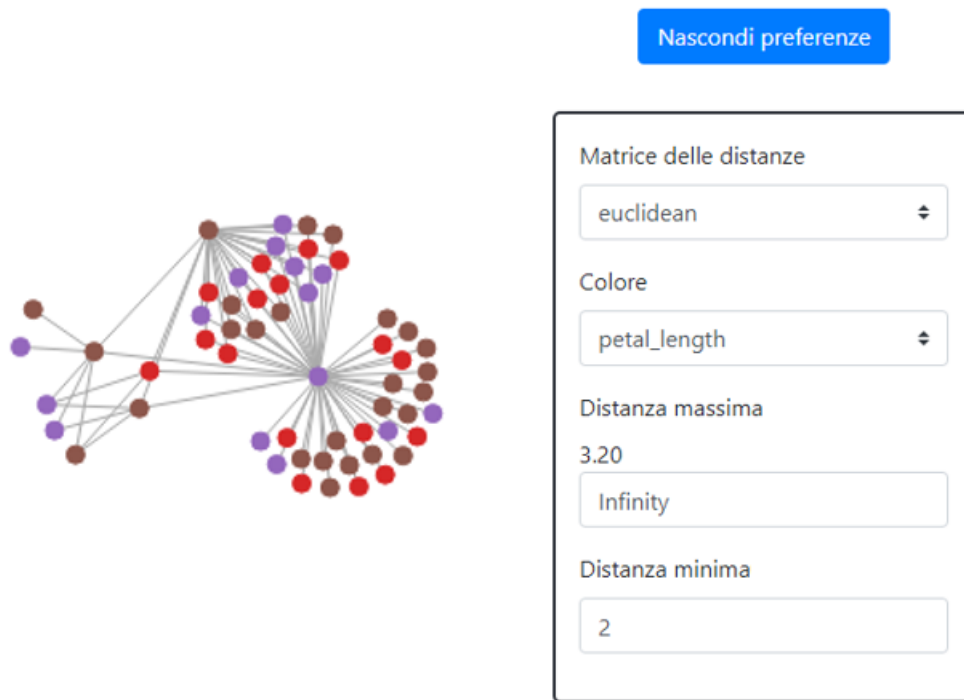


Figura 16: Form delle preferenze per il grafico Force Field ed esempio di visualizzazione

I nodi generati possono essere selezionati e spostati nello spazio tridimensionale per analizzare meglio il contenuto del grafico.

4.4.5 Proiezione Lineare Multi Asse

Per questa visualizzazione la form^G a destra permette di:

- Scegliere le dimensioni d'aggiungere alla visualizzazione;
- Scegliere la dimensione d'associare al colore dei punti.

Di seguito un esempio di visualizzazione dei dati con questo grafico e relativa form delle preferenze.

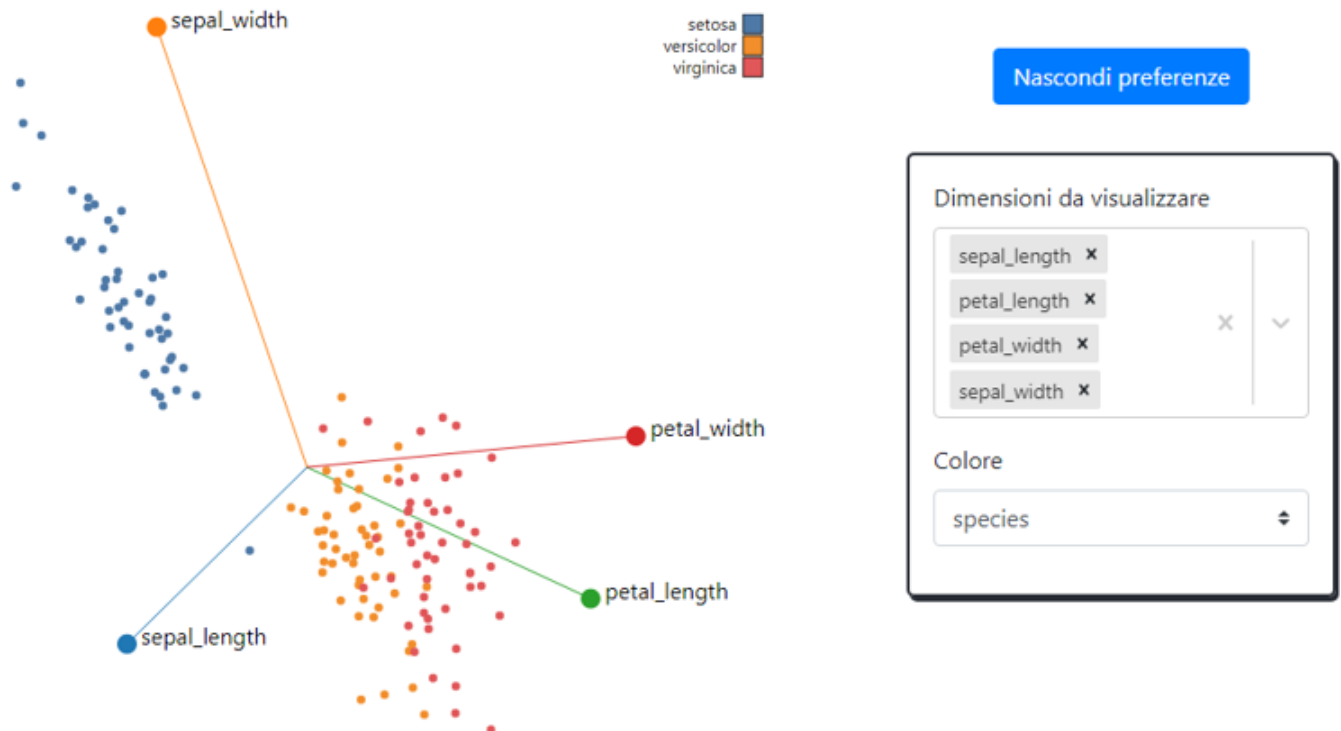


Figura 17: Form delle preferenze per il grafico PLMA ed esempio di visualizzazione

Come visibile in figura, in alto a destra sarà sempre disponibile una legenda dei colori per aiutare l'utente nell'analisi del grafico.

NB: la legenda sarà disponibile solo dopo aver associato una dimensione al colore dei punti dalla form delle preferenze.

Gli assi generati possono essere selezionati e spostati per muovere i punti nello spazio tridimensionale e analizzare meglio il contenuto del grafico.

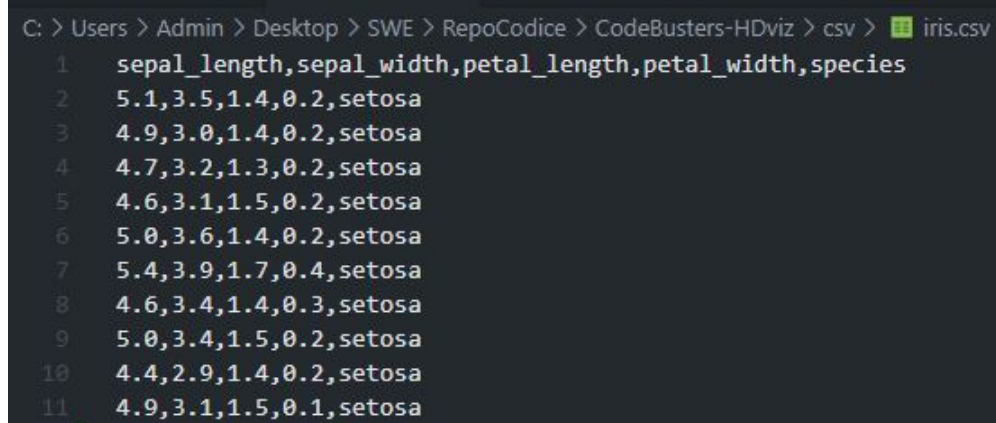
4.5 Formattazione dei file

Il file in formato CSV da dare in input all'applicazione dovrà seguire la formattazione standard per questo tipo di estensione. In particolare:

- La prima riga deve contenere il nome delle dimensioni, separato da una virgola l'una dall'altra;
- Le righe successive devono contenere i dati, rispettando l'ordine imposto dalla riga delle dimensioni e separando anch'essi con una virgola l'uno dall'altro;
- Al termine della riga andare a capo senza virgola.

NB: non è necessario inserire spazi.

Di seguito viene riportato un esempio di file CSV formattato correttamente.



```
C: > Users > Admin > Desktop > SWE > RepoCodice > CodeBusters-HDviz > csv > iris.csv
1  sepal_length,sepal_width,petal_length,petal_width,species
2  5.1,3.5,1.4,0.2,setosa
3  4.9,3.0,1.4,0.2,setosa
4  4.7,3.2,1.3,0.2,setosa
5  4.6,3.1,1.5,0.2,setosa
6  5.0,3.6,1.4,0.2,setosa
7  5.4,3.9,1.7,0.4,setosa
8  4.6,3.4,1.4,0.3,setosa
9  5.0,3.4,1.5,0.2,setosa
10 4.4,2.9,1.4,0.2,setosa
11 4.9,3.1,1.5,0.1,setosa
```

Figura 18: Esempio di file CSV correttamente formattato

5 Video Tutorial

Per guidare gli utenti sono stati realizzati dei video tutorial di pochi secondi per introdurre le funzionalità principali del prodotto.

5.1 Caricamenti dati da file CSV

[Visualizza il video tutorial su YouTube](#)

5.2 Caricamenti dati da Database

[Visualizza il video tutorial su YouTube](#)

5.3 Riduzione dimensionale

[Visualizza il video tutorial su YouTube](#)

5.4 Calcolo delle distanze

[Visualizza il video tutorial su YouTube](#)

5.5 Scatterplot Matrix

[Visualizza il video tutorial su YouTube](#)

5.6 Adjacency Matrix

[Visualizza il video tutorial su YouTube](#)

5.7 Heat Map

[Visualizza il video tutorial su YouTube](#)

5.8 Force Field

[Visualizza il video tutorial su YouTube](#)

5.9 Proiezione lineare Multi Asse

[Visualizza il video tutorial su YouTube](#)

6 Supporto tecnico

Nel caso venissero riscontrati malfunzionamenti, si prega di inviare una mail all'indirizzo:

codebusterswe@gmail.com

La mail deve preferibilmente seguire il seguente modello:

- **Oggetto:** [HDViz] "nome dell'evento da segnalare";
- **Corpo:**
 - Data in cui il malfunzionamento è stato riscontrato;
 - Descrizione del malfunzionamento;
 - Sistema operativo e browser^G in cui è avvenuto il problema.
- **Allegato:** è possibile allegare immagini per descrivere meglio il problema.

A Glossario

A

Adjacency matrix

La matrice delle adiacenze o matrice di connessione costituisce una particolare struttura dati comunemente utilizzata nella rappresentazione dei grafi finiti.

B

Browser

Applicazione per l'acquisizione, la presentazione e la navigazione di risorse sul web. Il programma implementa le funzionalità di client per il protocollo HTTP, che regola il download delle risorse dai server web, e quelle di visualizzazione dei contenuti ipertestuali e di riproduzione di contenuti multimediali.

C

Canberra

La distanza di Canberra è una misura numerica della distanza tra coppie di punti in uno spazio vettoriale.

Chebyshev

In matematica, la distanza di Čebyšëv, conosciuta anche come distanza della scacchiera o distanza di Lagrange, è una distanza su spazi vettoriali tale per cui la distanza tra due vettori è il valore massimo della loro differenza lungo gli assi.

Cluster

Un cluster è un insieme di oggetti che presentano tra loro delle similarità e, allo stesso modo, delle dissimilarità con oggetti in altri cluster.

CSV

Acronimo di Comma-Separated Values. Formato di file basato su file di testo utilizzato per l'importazione ed esportazione (ad esempio da fogli elettronici o database^G) di una tabella di dati.

D

D3.js

È una libreria JavaScript per creare visualizzazioni dinamiche ed interattive partendo da dati organizzati, visibili attraverso un comune browser^G.

Database

Letteralmente "base di dati". Rappresenta la versione digitale di un archivio di informazioni, ossia memorizza e organizza grandi moli di dati all'interno di dischi rigidi.

Dataset

Collezione di dati ordinati in tabella in cui ogni colonna rappresenta una dimensione e ogni riga un membro del dataset.

E

Euclidean

In matematica, la distanza euclidea è una distanza tra due punti, in particolare è una misura della lunghezza del segmento avente per estremi i due punti.

F

Fastmap

È un algoritmo di riduzione dimensionale per la visualizzazione di dataset^G tradizionali e multimediali. Si basa sulla mappatura di oggetti in punti in uno spazio k-dimensionale.

Force field

Grafico che traduce le distanze nello spazio a molte dimensioni in forze di attrazione e repulsione tra i punti proiettati nello spazio bidimensionale (o anche tridimensionale). Questo grafico esegue una riduzione dimensionale preservando, o addirittura evidenziando, le strutture presenti nei dati.

Form

Il termine form indica la parte di interfaccia utente di un'applicazione che consente all'utente di inviare uno o più dati liberamente inseriti dallo stesso. Per descriverlo può essere utile la metafora della "scheda da compilare" per l'inserimento di dati.

H

Heatmap

Grafico che trasforma la distanza tra i punti in colori più o meno intensi, permettendo così di comprendere quali oggetti sono più vicini tra loro. Per una buona visualizzazione è utile accompagnarlo con l'ordinamento dei dati in modo che le strutture presenti siano evidenziate. Questo grafico e la relativa operazione di ordinamento sono reperibili nella libreria D3.js^G.

I

Isomap

È un efficiente algoritmo non lineare per la riduzione dimensionale. L'algoritmo offre un semplice metodo per stimare un manifold^G basandosi su una stima dei neighbors^G di ciascun punto.

L

LLE

Acronimo di Locally-Linear Embedding. È un algoritmo di riduzione dimensionale non lineare. Si basa sui più vicini neighbors^G di ciascun punto e da un'ottimizzazione ad autovettori.

M

Manhattan

In matematica, distanza di Manhattan (chiamata anche geometria del taxi) è un concetto geometrico secondo il quale la distanza tra due punti è la somma del valore assoluto delle differenze delle loro coordinate.

Manifold

Detta anche varietà differenziabile, è una struttura matematica su cui si basano alcuni metodi di riduzione della dimensionalità. Si tratta di uno spazio matematico dove localmente viene a ricrearsi uno spazio euclideo (di una specifica dimensione).

N

Neighbors

Sono il numero di vicini che vengono determinati sul manifold^G in base alla distanza Euclidea. Possono essere scelti per gli algoritmi IsoMap^G ed LLE^G.

P

PLMA

Acronimo di Proiezione Lineare Multi Asse. Grafico che posiziona i punti dello spazio multidimensionale in un piano cartesiano (con assi "dragabili"), effettuando quindi sui dati una riduzione a due dimensioni.

R

Repository

Ambiente di un sistema informativo, in cui vengono gestiti i metadati, attraverso tabelle relazionali; l'insieme di tabelle, regole e motori di calcolo tramite cui si gestiscono i metadati prende il nome di metabase.

S

Scatter plot

Grafico di dispersione a due dimensioni. I dati sono riportati su uno spazio cartesiano: una dimensione sull'asse delle ascisse e una su quello delle ordinate.

Scatter plot matrix

Grafico formato da più scatter plot^G disposti a matrice, dove in ognuno vengono messe in relazione due dimensioni diverse. Esso aiuta l'analista a trovare dimensioni con forti correlazioni e più dimensioni che danno la stessa informazione. Il grafico è reperibile nella libreria D3.js^G.

T

TSNE

Acronimo di T-Distributed Stochastic Neighbor Embedding. È un algoritmo di riduzione dimensionale non lineare che si presta particolarmente all'embedding di dataset^G ad alta dimensionalità in uno spazio a due o tre dimensioni, nel quale possono essere visualizzati tramite un grafico di dispersione. L'algoritmo modella i punti in modo che oggetti vicini nello spazio originale risultino vicini nello spazio a dimensionalità ridotta e oggetti lontani risultino lontani, cercando di preservare la struttura locale.