

C~~ODE~~BUSTERS

Progetto: *HD Viz*
codebusterswe@gmail.com

Manuale Utente

Informazioni sul documento

Versione	2.0.0-1.8
Approvatori	Scialpi Paolo
Redattori	Sassaro Giacomo Baldisseri Michele
Verificatori	Pirola Alessandro Zenere Marco
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
2.0.0-1.8	15-05-2021	Scialpi Paolo	Responsabile	Approvazione del documento
1.1.0-1.8	14-05-2021	Zenere Marco	Verificatore	Revisione complessiva del documento
1.0.3-1.8	10-05-2021	Baldisseri Michele, Pirolo Alessandro	Amministratore, Verificatore	Aggiornata §5 e verifica
1.0.2-1.8	06-05-2021	Sassaro Giacomo, Pirolo Alessandro	Programmatore, Verificatore	Estensione §A e verifica
1.0.1-1.8	05-05-2021	Sassaro Giacomo, Pirolo Alessandro	Programmatore, Verificatore	Aggiornata §4 e verifica
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 §3, inserite immagini in §4, stesura §6 e verifica
0.0.2-0.4	20-03-2021	Sassaro Giacomo, Pirolo Alessandro	Amministratore, Verificatore	Stesura di §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	6
1.1	Scopo del documento	6
1.2	Scopo del prodotto	6
1.3	Glossario	6
2	Requisiti minimi di sistema	7
2.1	Requisiti di sistema	7
2.2	Requisiti hardware	7
2.3	Browser	8
3	Installazione	9
3.1	Clonare la repository	9
3.2	Importare il database	9
3.3	Avviare il server	9
3.4	Avviare la web app	10
4	Istruzioni all'uso	11
4.1	Ripristino sessione	11
4.2	Caricamento dati	12
4.2.1	Caricamento dati dal database	13
4.2.2	Caricamento dati da CSV	14
4.3	Riduzione dimensionale	15
4.4	Calcolo della distanza	16
4.5	Scelta del grafico e visualizzazione dei dati	17
4.5.1	Scatterplot matrix	18
4.5.2	Adjacency matrix	20
4.5.3	Heat map	22
4.5.4	Force field	23
4.5.5	Proiezione lineare multi asse	24
4.6	Esportazione della sessione	25
4.7	Formattazione dei file	26
5	Video tutorial	27
5.1	Salvataggio e caricamento sessione	27
5.2	Caricamenti dati da file CSV	27
5.3	Caricamenti dati da database	27
5.4	Riduzione dimensionale	27
5.5	Calcolo delle distanze	27
5.6	Scatterplot matrix	27
5.7	Adjacency matrix	27
5.8	Heat map	27
5.9	Force field	27
5.10	Proiezione lineare multi asse	27

6	Supporto tecnico	28
A	Glossario	29

Elenco delle tabelle

1	Requisiti di sistema	7
2	Requisiti hardware	7
3	Prestazioni ottimali	7
4	Browser e versioni compatibili	8

Elenco delle figure

1	Avvio dell'applicazione	10
2	Sezione per il ripristino della sessione	11
3	Messaggi d'esito del caricamento dati nel sistema	12
4	Caricamento dati da database	13
5	Caricamento dati da file CSV	14
6	Finestra per la riduzione dimensionale tramite algoritmo	15
7	Finestra per la riduzione dimensionale tramite il calcolo delle distanze	16
8	Form delle preferenze per il grafico Scatterplot Matrix	18
9	Esempio di visualizzazione dei dati con il grafico Scatterplot Matrix	19
10	Form delle preferenze per il grafico Adjacency Matrix	20
11	Esempio di visualizzazione dei dati con il grafico Adjacency Matrix	21
12	Form delle preferenze per il grafico Heat Map	22
13	Esempio di visualizzazione dei dati con il grafico Heat Map	22
14	Form delle preferenze per il grafico Force Field ed esempio di visualizzazione	23
15	Form delle preferenze per il grafico PLMA ed esempio di visualizzazione	24
16	Sezione per l'esportazione della sessione	25
17	Esempio di file CSV correttamente formattato	26

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 *HD Viz* 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

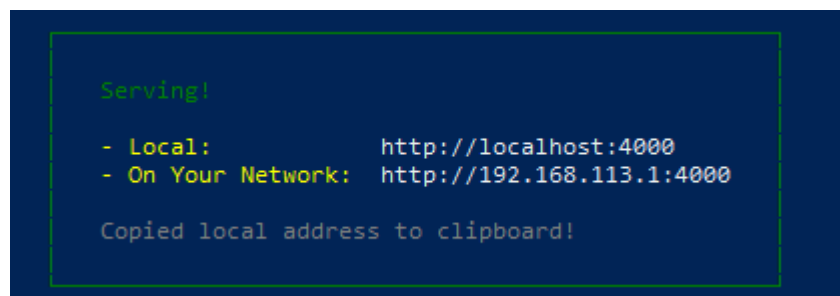
4) nel caso di primo avvio dell'applicazione sul sistema, digitare:

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.

Al primo avvio sono selezionabili solamente le voci che permettono di caricare dei dati o recuperare una sessione di lavoro. Le voci rimanenti sono disabilitate in quanto necessitano dati da elaborare. Questo è visibile all'utente ed è segnalato con un tooltip passando sopra alle varie voci del menù con il cursore.

4.1 Ripristino sessione

Cliccando sul pulsante *Carica/Salva Sessione* un utente che aveva precedentemente utilizzato il sito ed esportato una sessione, come spiegato a §4.6, può riportare lo stato dell'applicazione a come lo aveva lasciato.

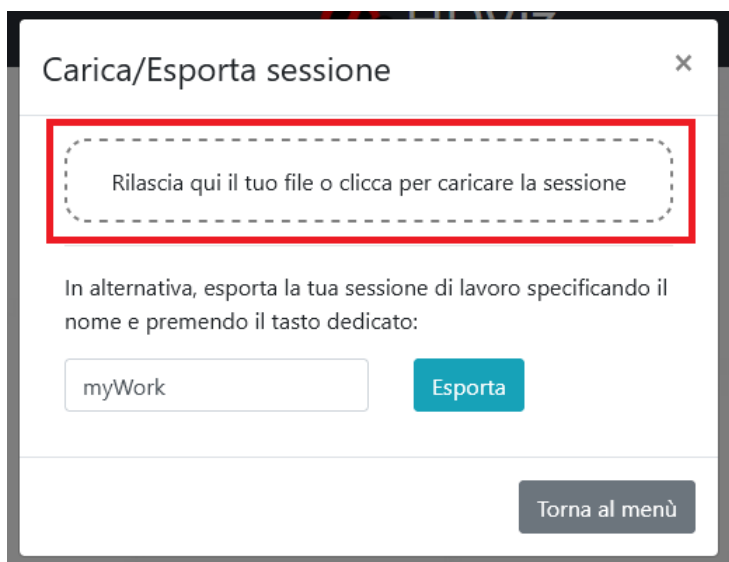


Figura 2: Sezione per il ripristino della sessione

Per ripristinare la sessione sarà sufficiente trascinare il file JSON contenente la sessione nel riquadro evidenziato in Figura 2, oppure cliccare su tale riquadro e successivamente selezionare il file JSON dal proprio dispositivo. Al termine del ripristino viene notificato il successo dell'operazione o l'eventuale fallimento nel caso il file non sia valido. Il ripristino della sessione andrà a caricare dimensioni e matrici delle distanze, ottenute tramite riduzioni dimensionali, e tutte le preferenze di visualizzazione nei vari grafici che l'utente aveva utilizzato durante la propria sessione di lavoro. Anche i dati originali sono interessati dal processo di ripristino e non risulta quindi necessario svolgere le azione illustrate a §4.2.

4.2 Caricamento dati

Tra le voci disponibili al primo avvio dell'applicazione ne sono presenti due che possono essere utilizzate per caricare i dati nel sistema:

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

Entrambe aprono una finestra di dialogo per compiere le operazioni indicate. 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 delle finestre.

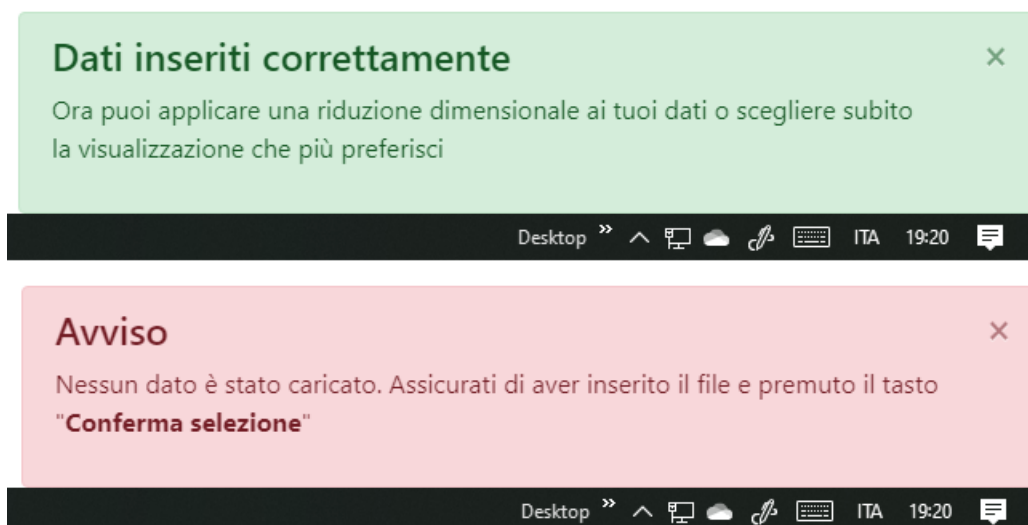


Figura 3: Messaggi d'esito del caricamento dati nel sistema

4.2.1 Caricamento dati dal database

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.

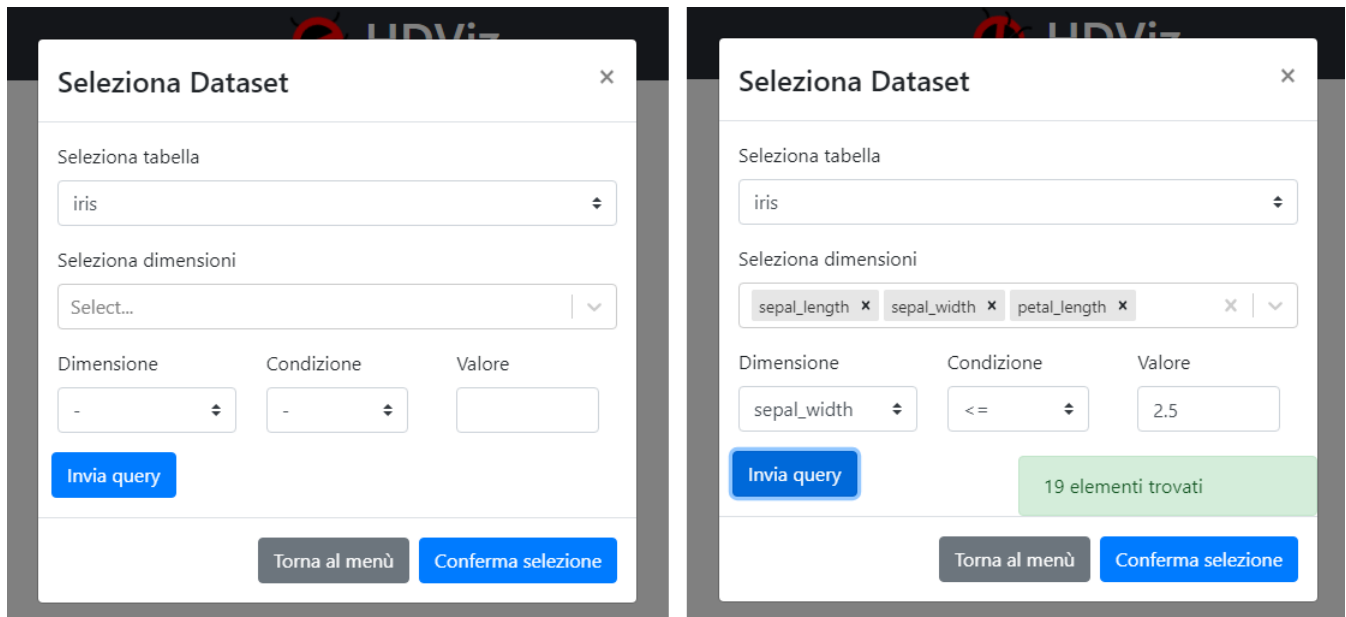


Figura 4: Caricamento dati da database

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.

4.2.2 Caricamento dati da CSV

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'esplorazione 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.

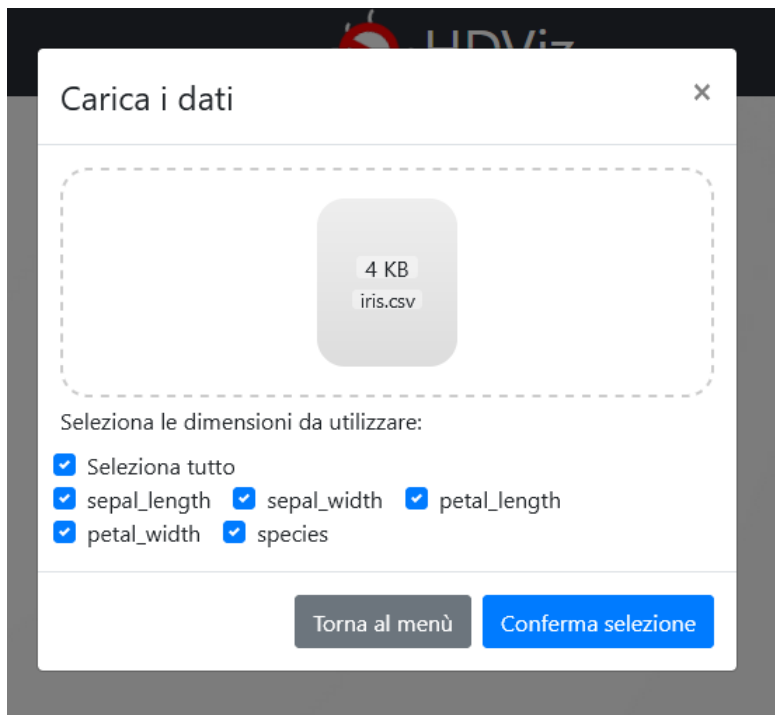


Figura 5: Caricamento dati da file CSV

4.3 Riduzione dimensionale

In seguito al caricamento dei dati vengono rese disponibili ulteriori voci, tra cui ”*Riduci dimensioni*”. Cliccando su questa voce si apre una finestra che permette di scegliere quali dimensioni saranno interessate dal processo, se normalizzare o meno i dati e quale algoritmo utilizzare tra:

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

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

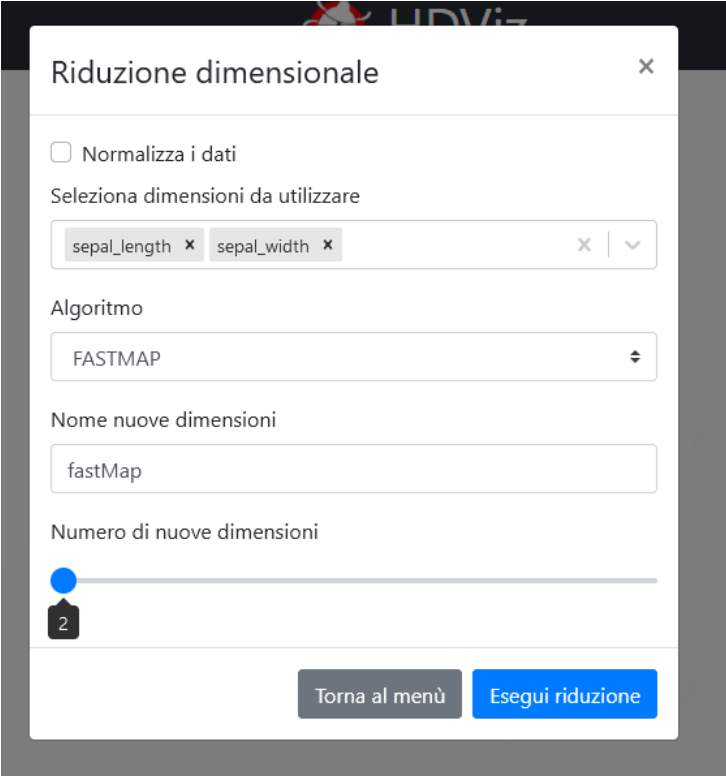


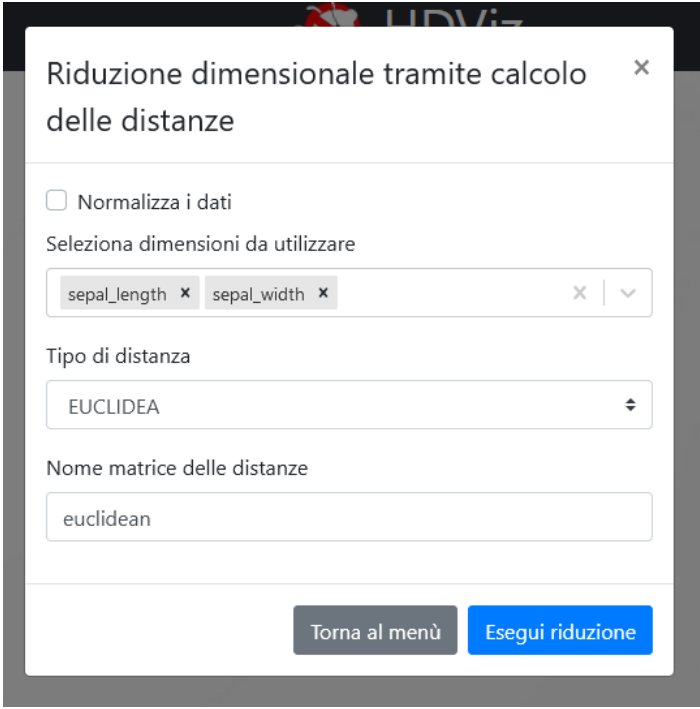
Figura 6: Finestra per la riduzione dimensionale tramite algoritmo

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

4.4 Calcolo della distanza

In seguito al caricamento dei dati vengono rese disponibili ulteriori voci, tra cui ”*Calcola distanza*”. Cliccando su questa voce si apre una finestra che permette di scegliere quali dimensioni utilizzare per creare una matrice delle distanze, quale nome assegnarvi, se normalizzare o meno i dati e quale funzione di distanza usare tra:

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



The image shows a software window titled "Riduzione dimensionale tramite calcolo delle distanze" with a close button (X) in the top right corner. Inside the window, there is a checkbox labeled "Normalizza i dati" which is currently unchecked. Below this is a label "Seleziona dimensioni da utilizzare" followed by a container showing two selected dimensions: "sepal_length" and "sepal_width", each with a small 'x' icon to its left and a dropdown arrow to its right. Underneath is a label "Tipo di distanza" followed by a dropdown menu currently set to "EUCLIDEA". Below that is a label "Nome matrice delle distanze" followed by a text input field containing the word "euclidean". At the bottom of the window, there are two buttons: "Torna al menù" (grey) and "Esegui riduzione" (blue).

Figura 7: 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).

4.5 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;
- Adjency matrix^G;
- Heat map^G;
- Force field^G;
- PLMA (Proiezione lineare multi asse)^G.

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 impostati a " *Nessuna dimensione*" quindi è corretto che il grafico non sia già visualizzato. 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 centrare il grafico nello schermo per concentrarsi solamente sull'analisi del grafico.

4.5.1 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

Asse uno

sepal_length

Asse due

sepal_width

Asse tre

petal_length

Asse quattro

petal_width

Asse cinque

Nessuna dimensione

Colore

species

Figura 8: Form delle preferenze per il grafico Scatterplot Matrix

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

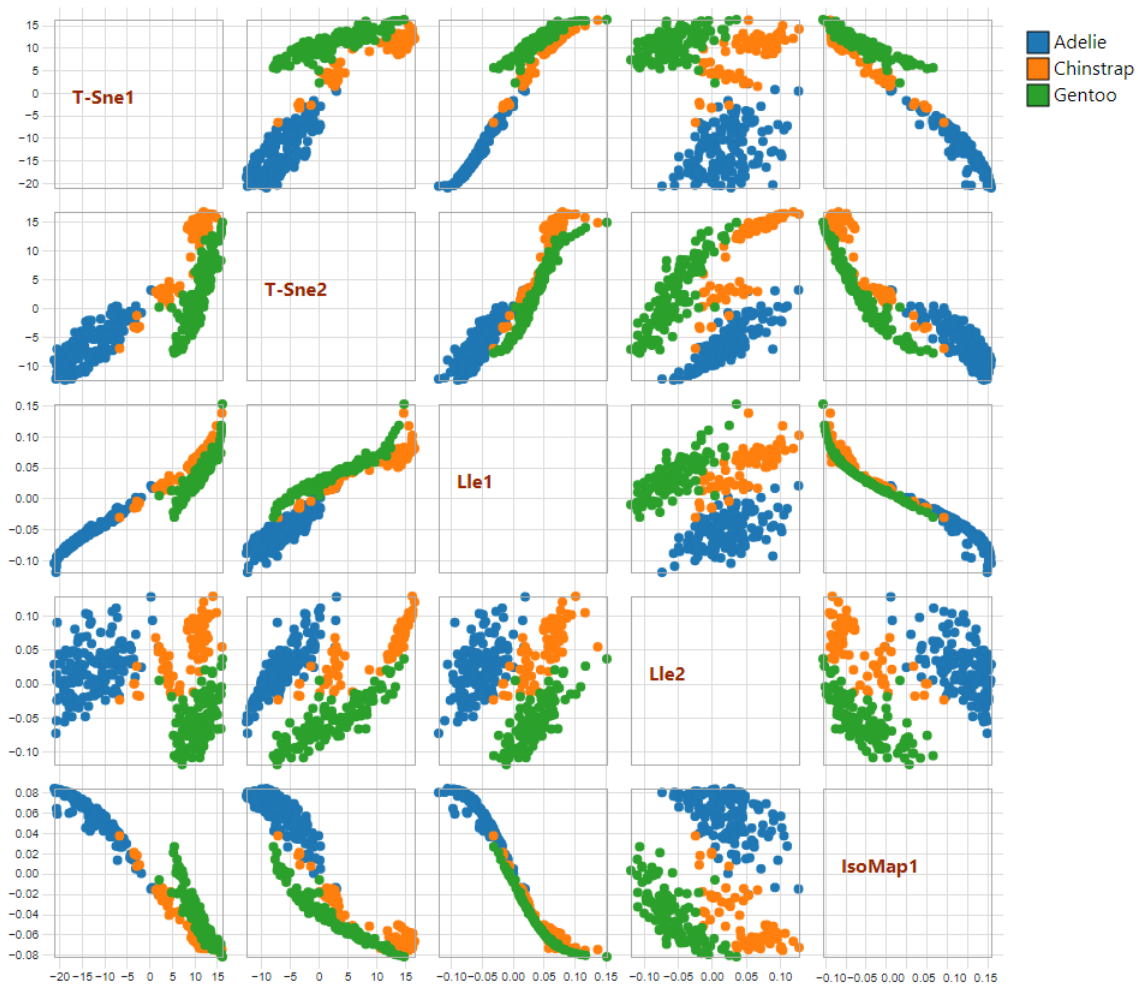


Figura 9: Esempio di visualizzazione dei dati con il grafico Scatterplot Matrix

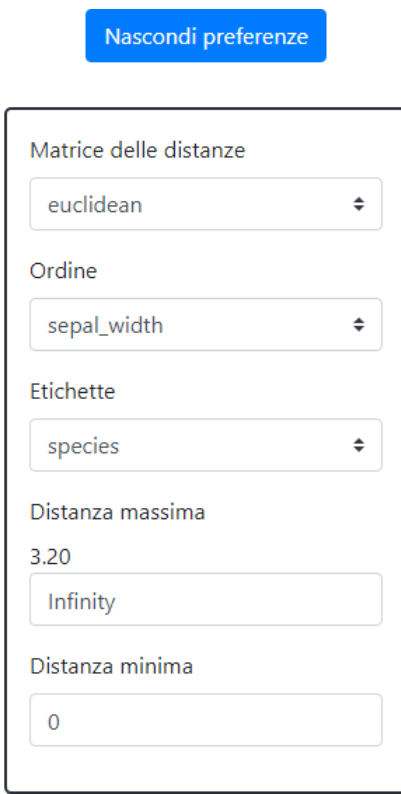
Come visibile in figura, in alto a destra sarà sempre disponibile una legenda dei colori per aiutare l'utente nell'analisi del grafico.

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

4.5.2 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 black 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 10: Form delle preferenze per il grafico Adjacency Matrix

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

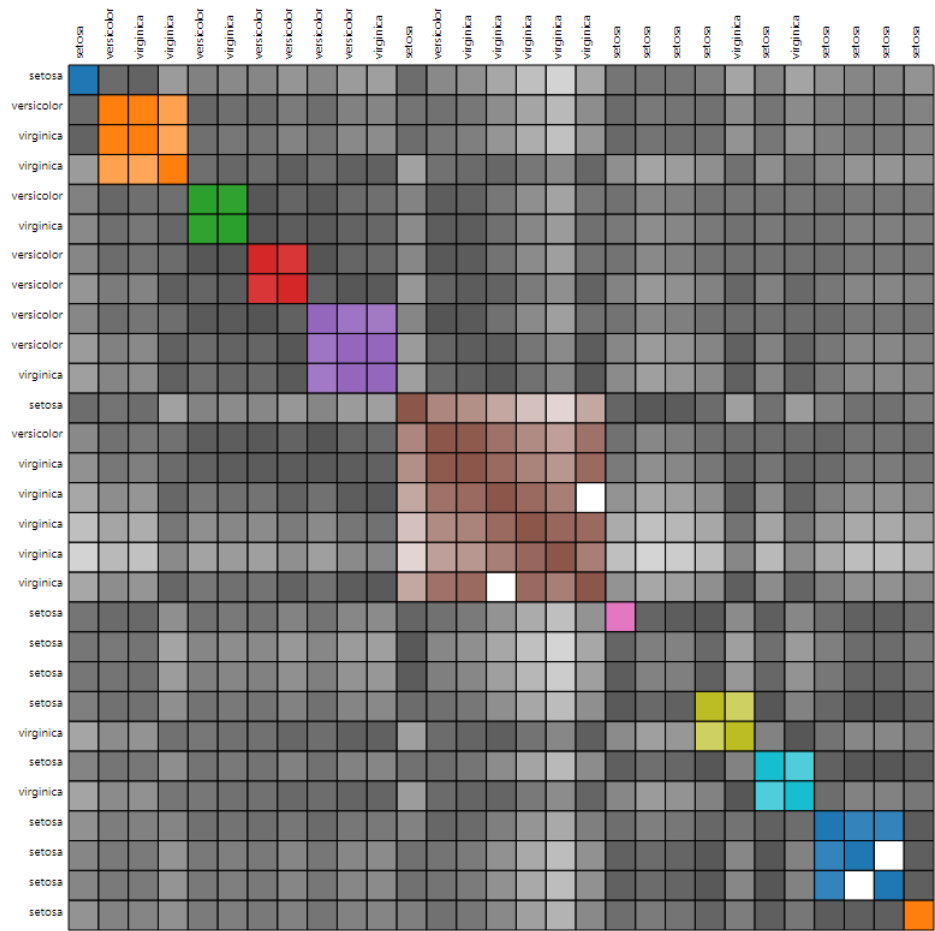


Figura 11: Esempio di visualizzazione dei dati con il grafico Adjacency Matrix

4.5.3 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 12: Form delle preferenze per il grafico Heat Map

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

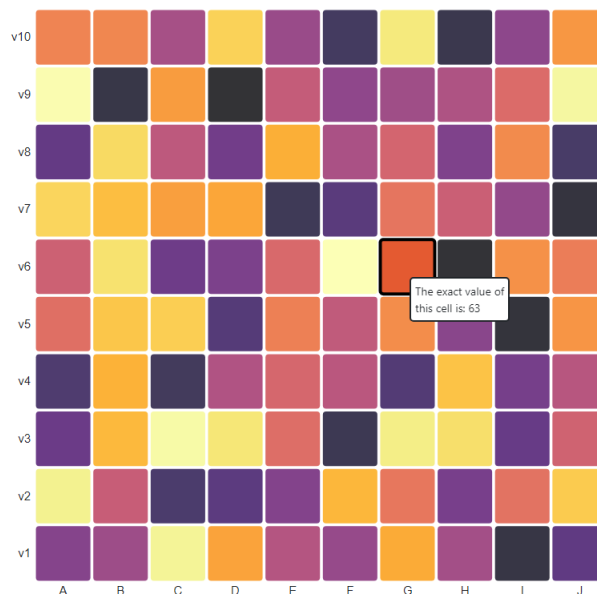


Figura 13: Esempio di visualizzazione dei dati con il grafico Heat Map

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.

4.5.4 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.

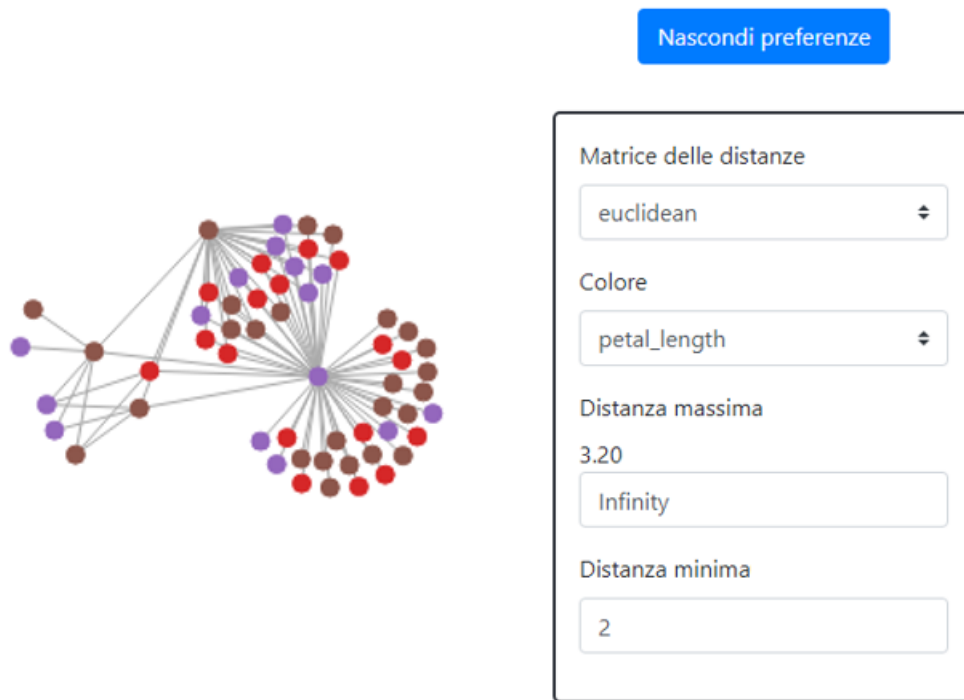


Figura 14: Form delle preferenze per il grafico Force Field ed esempio di visualizzazione

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

4.5.5 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.

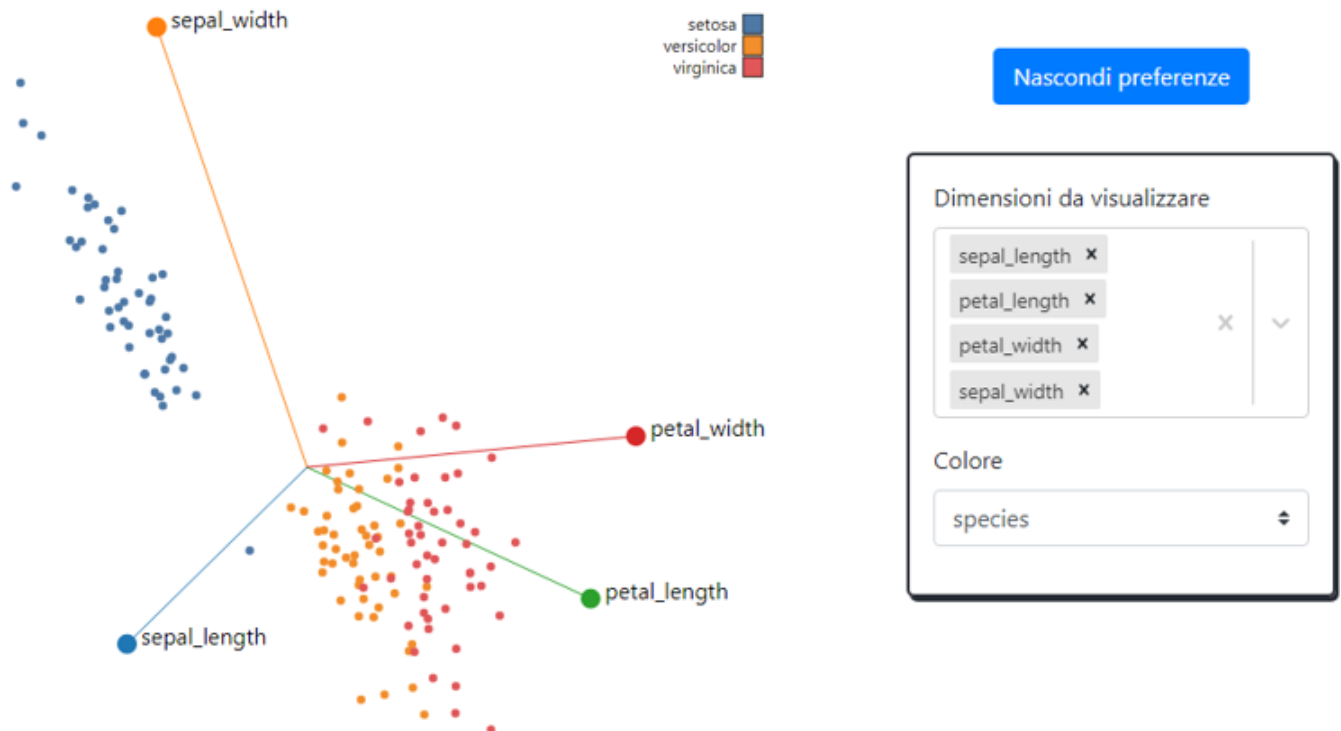


Figura 15: Form delle preferenze per il grafico PLMA ed esempio di visualizzazione

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

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.

4.6 Esportazione della sessione

Nel caso in cui l'utente desidera salvare lo stato dell'applicazione, nel menu è disponibile la voce *Carica/Salva sessione*. L'applicazione permette di esportare lo stato attuale tramite il download di un file in formato JSON che potrà successivamente essere ripristinato (vedere §4.1).



Figura 16: Sezione per l'esportazione della sessione

Per esportare la sessione di lavoro l'utente può associare un nome al file e successivamente premere sul pulsante *Esporta*. Il download partirà automaticamente e il file sarà disponibile nel dispositivo utilizzato.

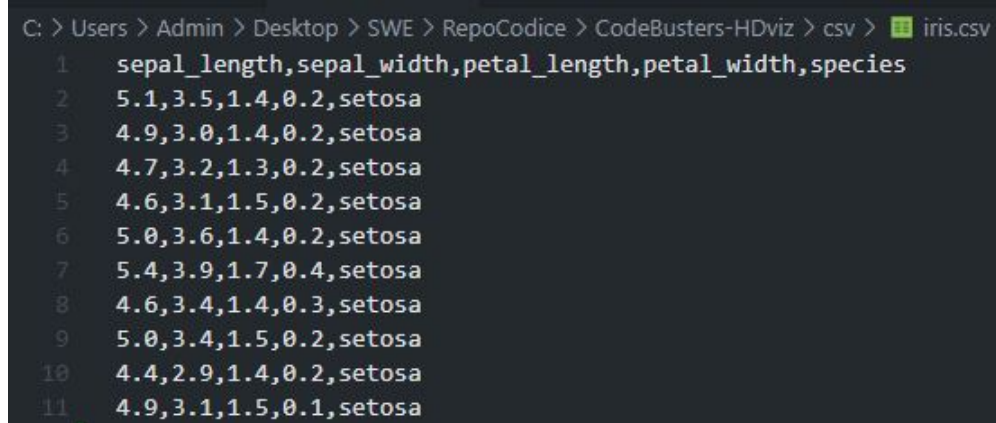
4.7 Formattazione dei file

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 17: 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 Salvataggio e caricamento sessione

[Visualizza il video tutorial su YouTube](#)

5.2 Caricamenti dati da file CSV

[Visualizza il video tutorial su YouTube](#)

5.3 Caricamenti dati da database

[Visualizza il video tutorial su YouTube](#)

5.4 Riduzione dimensionale

[Visualizza il video tutorial su YouTube](#)

5.5 Calcolo delle distanze

[Visualizza il video tutorial su YouTube](#)

5.6 Scatterplot matrix

[Visualizza il video tutorial su YouTube](#)

5.7 Adjacency matrix

[Visualizza il video tutorial su YouTube](#)

5.8 Heat map

[Visualizza il video tutorial su YouTube](#)

5.9 Force field

[Visualizza il video tutorial su YouTube](#)

5.10 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

PCA

Acronimo di Principal Component Analysis. È una tecnica per la semplificazione dei dati utilizzata nell'ambito della statistica multivariata. Lo scopo della tecnica è quello di ridurre il numero più o meno elevato di variabili che descrivono un insieme di dati a un numero minore di variabili latenti, limitando il più possibile la perdita di informazioni.

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.

U

UMAP

Algoritmo di riduzione dimensionale per aumentare le prestazioni del clustering basato sulla densità. Come t-SNE^G, non preserva completamente la densità.