

Progetto A_Clus - Documentazione

A cura di: Vito Stefano
Lorenzo Gelao

3 aprile 2025

Indice

1	Introduzione	3
1.1	Agglomerative Clustering	3
1.1.1	Framework Analitico	3
2	Guida all'installazione	4
2.1	Installazione JDK	4
2.2	Variabili di ambiente	4
2.3	Installazione MySQL	5
3	Eseguire A-CLus Base	8
3.1	Interfaccia iniziale	9
3.2	Selezione delle operazioni disponibili	10
3.3	Percorso operativo - Opzione 2 (Generazione nuovo dendrogramma)	10
3.4	Inserimento Profondità	11
3.4.1	Elaborazione con Single-link	11
3.5	Inserimento nome file	11
3.5.1	Elaborazione con Average-link	12
3.6	Percorso operativo - Opzione 1 (Caricamento dendrogramma esi- stente)	12
3.7	Inserimento nome archivio con estensione	13
4	TEST A-CLUS BASE	13
4.1	Gestione di tabelle inesistenti	13
4.2	Validazione delle selezioni di menù	14
4.2.1	Controllo sull'esistenza degli archivi	14
4.2.2	Validazione parametri di profondità	14
4.2.3	Controllo selezioni metodologiche	15
4.2.4	Verifica di compatibilità dimensionale	15
4.3	Gestione connettività client-server	15
4.3.1	Elaborazione su dataset vuoti	16
5	Eseguire A-CLus Esteso	16

1 Introduzione

Il progetto in questione verte sull'argomento dell'*Agglomerative Clustering*, una tecnica di clusterizzazione basata sui metodi Single distance e Average Distance.

Il progetto in questione è suddiviso in una parte client e una server, che comunicando tra loro, generano il dendrogramma, permettendo inoltre di visualizzare e memorizzare tali risultati o di caricarne dei precedenti.

È inoltre possibile visualizzare nuovamente dei file caricati in passato per visualizzare i cluster e i dendrogrammi associati.

1.1 Agglomerative Clustering

L'algoritmo di clustering utilizzato in tale progetto, come si può desumere dal nome, sfrutta il concetto di clustering agglomerativo. In pratica, tratta ciascun cluster in maniera separata dagli altri, unendo progressivamente quelli più vicini, in base a due criteri principali, nel nostro caso.

Il principale vantaggio rispetto ad altri algoritmi di clustering, come il k-means, è che in questo modo non è necessario specificare in anticipo la quantità di cluster da analizzare.

I criteri utilizzati nel progetto A-CLus sono i seguenti:

1. **Single-Link:** tale criterio determina la distanza minima tra i punti dei vari cluster

$$D(C1, C2) = \min_{(t1 \in C1, t2 \in C2)} (dist(t1, t2)) \quad (1)$$

Durante l'anno accademico 2024/2025, l'oggetto di ricerca è stato incentrato su "H-CLUS", una piattaforma con architettura client-server dedicata all'analisi dei dati mediante algoritmi di clustering gerarchico agglomerativo. La componente server esegue le operazioni di clustering impiegando le metodologie Single Link o Average Link per il calcolo delle distanze inter-cluster e la successiva costruzione del dendrogramma. L'applicativo client, implementato in linguaggio Java, offre agli utenti diverse funzionalità: il caricamento o la creazione di istanze HierarchicalClusterMiner, la rappresentazione grafica del dendrogramma e l'archiviazione dei risultati per analisi successive. È inoltre disponibile la funzione di importazione di file precedentemente salvati, consentendo agli utenti di riesaminare i cluster e i relativi dendrogrammi.

1.1.1 Framework Analitico

La tecnica di clustering agglomerativo implementata nel sistema "A-CLUS" costituisce una metodologia avanzata per l'identificazione di correlazioni latenti nei dataset.

Diversamente da tecniche alternative come il k-means, l'approccio agglomerativo elimina la necessità di predefinire il numero di raggruppamenti. La procedura inizializza ciascun elemento come cluster individuale e procede con l'unificazione sequenziale dei cluster con maggiore affinità, applicando metodologie quali:

Con il procedere delle aggregazioni tra cluster, si sviluppa un dendrogramma che illustra la gerarchia delle aggregazioni. Il procedimento continua fino al raggiungimento di un cluster unificato o fino a una soglia di profondità stabilita dall'utilizzatore. La visualizzazione mediante dendrogramma rende la metodologia particolarmente comprensibile, facilitando l'esplorazione strutturale dei dati a diversi livelli di granularità. Inoltre, questa strategia non è influenzata dalla configurazione iniziale dei centroidi, riducendo così la probabilità di risultati inconsistenti e garantendo una rappresentazione più accurata dell'organizzazione interna del dataset. RiprovaClaude può commettere errori. Verifica sempre le risposte con attenzione.

2 Guida all'installazione

Prima di essere in grado di eseguire il programma, è necessario eseguire il file `risorse.bat` contenuto nella cartella Risorse.

Una volta eseguito il file, si aprirà una pagina di Powershell e seguirà un download.



Figura 1: Schermata di download

Al termine del download del file compresso delle risorse, verranno estratti i file necessari per l'esecuzione del programma.

Per il progetto in questione è necessario installare il Java Developer Kit (JDK) nella versione 22.0.1 e il software di gestione del database MySQL nella sua versione 8.0.39.

La prima scheda di installazione che apparirà è quella del JDK.

Verrà chiesto se si vuole eseguire l'installazione del JDK tramite permessi di amministratore, cliccare su Sì.

2.1 Installazione JDK

Una volta confermata l'esecuzione come amministratore, si procederà all'installazione del JDK.

Cliccare su **Avanti**, nuovamente **Avanti** e infine, nel caso in cui l'installazione sia andata a buon fine, si dovrà cliccare sul tasto **Chiudi**.

2.2 Variabili di ambiente

è consigliato creare una variabile di ambiente per il JDK. Per farlo, si dovrà cercare la voce **variabili di ambiente** nella barra di ricerca di Windows e cliccare su Modifica le variabili di ambiente per il tuo account.

Successivamente si dovrà cliccare sul pulsante **Variabili di ambiente** e si aprirà una schermata con le variabili di ambiente.

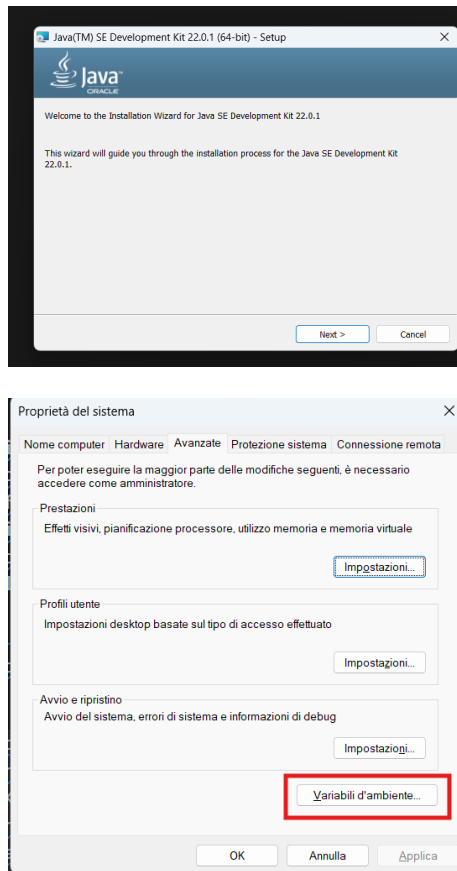


Figura 2: Schermata delle variabili di ambiente

Nella schermata, si dovrà cliccare sulla variabile **PATH** e modificarla in modo da aggiungere il percorso del JDK, inserendo il percorso in cui è stato installato il JDK.

Una volta completata l'installazione del JDK, si procederà automaticamente con l'installazione del MySQL.

2.3 Installazione MySQL

Anche qui sarà necessario fornire i permessi di amministratore, cliccare dunque su Sì.

Schermata iniziale

Dopodiché si aprirà la schermata di installazione di MySQL.

Selezionare il tipo di setup **Full** e cliccare su **Next**. Tale scelta permetterà di installare tutti i componenti necessari per l'esecuzione del programma.

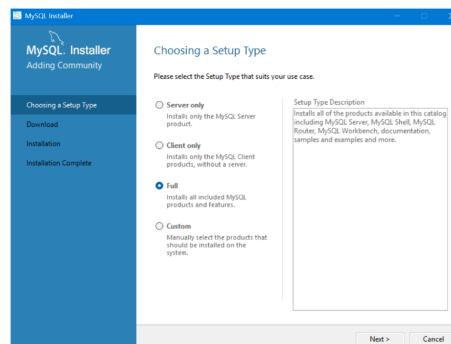


Figura 3: Scelta del setup

Download dei file

Si passerà alla schermata di download dei file necessari per l'installazione, cliccare su **Execute**.

AVVERTENZA

Tale schermata potrebbe richiedere un po' di tempo per il download dei file, a seconda della velocità della connessione internet.

Dopo aver scaricato i file, si dovrà cliccare nuovamente su **Execute**.

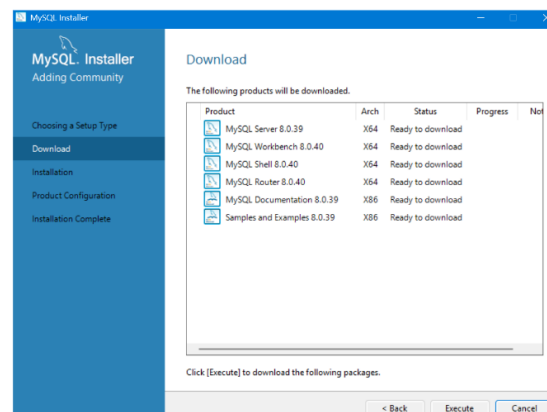


Figura 4: Download MySQL

Configurazione del account MySQL

Una volta terminato il download, si procederà con la configurazione dell'account MySQL. Si dovrà quindi scegliere una password, necessaria per accedere al database.

Si giungerà infine alla schermata di applicazione della configurazione del database. Cliccare su **Execute** per applicare la configurazione.

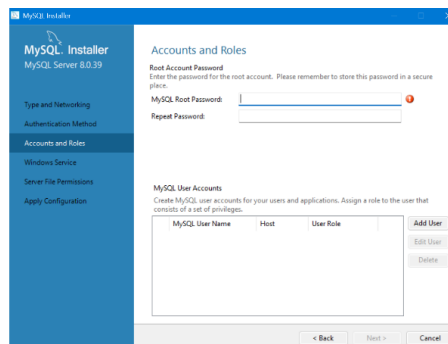


Figura 5: Configurazione account MySQL

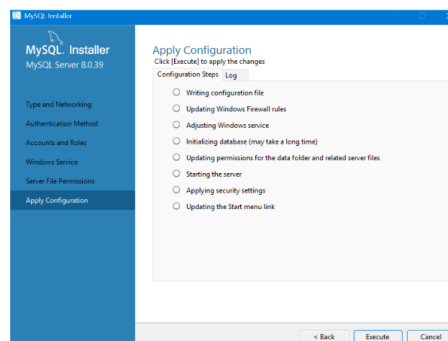


Figura 6: Completamento installazione

Una volta terminata anche l'installazione del MySQL, si dovrà cliccare su **Finish** per chiudere l'installler.

AVVERTENZA

Per un corretto funzionamento sia del JDK che del MySQL, è consigliato riavviare il computer.

Schermata finale

Al termine dell'installazione il programma `resources.bat` mostrerà la seguente schermata.



Figura 7: Completamento installazione

3 Eseguire A-CLus Base

Operazioni preliminari

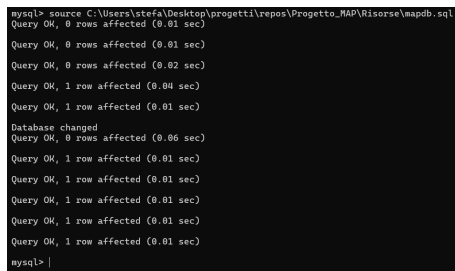
Prima di poter eseguire il programma, occorre importare il database di A-CLus in MySQL.

Tale file è presente nella cartella **A-CLus/Risorse** e si chiama **mapdb.sql**.

Per importare il database, è necessario semplicemente accedere al MySQL tramite root e digitare il seguente comando **source** seguito dal percorso del file **mapdb.sql**.

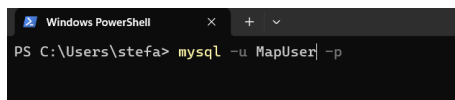
Tale percorso si può ottenere semplicemente trascinando il file del database sul terminale di MySQL.

Alle fine dell'importazione, il terminale di MySQL dovrebbe mostrare un messaggio simile al seguente:



```
mysql> source C:\Users\stefa\Desktop\progetti\repos\Progetto_MAP\Risorse\mapdb.sql
Query OK, 0 rows affected (0.01 sec)
Query OK, 0 rows affected (0.02 sec)
Query OK, 1 row affected (0.04 sec)
Query OK, 1 row affected (0.01 sec)
Database changed
Query OK, 0 rows affected (0.06 sec)
Query OK, 1 row affected (0.01 sec)
Query OK, 1 row affected (0.01 sec)
Query OK, 1 row affected (0.01 sec)
Query OK, 1 row affected (0.01 sec)
Query OK, 1 row affected (0.01 sec)
mysql>
```

È possibile verificare la corretta importazione anche uscendo da MySQL e accedere con le credenziali dell'utente MapUser, come mostrato nella figura seguente:



```
Windows PowerShell
PS C:\Users\stefa> mysql -u MapUser -p
```

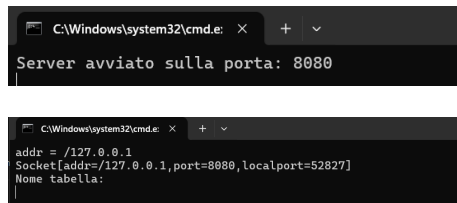
La password dell'utente MapUser è **map**. Se viene confermato l'accesso al database, significa che l'importazione è avvenuta con successo.

Sarà sufficiente rieffettuare l'importazione del database nel nuovo utente MapUser.

Avvio dell'applicazione

Per inizializzare correttamente la versione standard di A-CLus, attenersi alla seguente sequenza operativa:

1. Accedere alla directory **A-CLus_Base/Bat**
2. Eseguire il file **Start_Server.bat** facendo un doppio click su di esso, assicurandosi che venga mostrato il seguente messaggio a finestra
3. Dopo aver verificato il corretto funzionamento del server, eseguire il file **Start_Client.bat** facendo un doppio click su di esso; se non ci sono problemi con il server, apparirà la seguente schermata:



```
C:\Windows\system32\cmd.e: x + v
Server avviato sulla porta: 8080

C:\Windows\system32\cmd.e: x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52827]
Nome tabella:
```

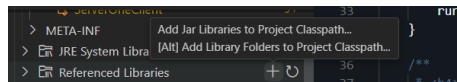
Avvertenza

È necessario mantenere attivo il terminale del server per garantire la comunicazione tra le componenti client e server.

Avvio dell'ambiente di sviluppo

Per aprire correttamente il codice sorgente di A-CLus, è necessario seguire le seguenti operazioni nell'ordine in cui sono presentate:

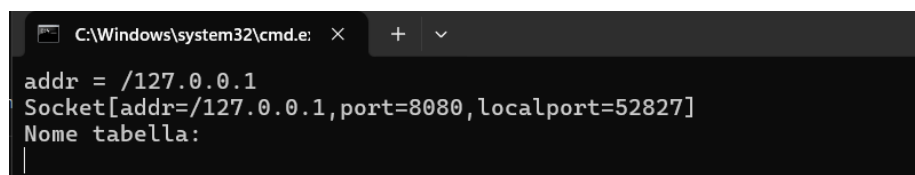
1. Aprire la cartella del progetto **A-Clus_base**
2. Incorporare nelle librerie del progetto **Server** il file **mysql.connector.jar** presente nella cartella **A-CLus/Risorse**. Nel caso si stia eseguendo il progetto in Visual Studio Code, è possibile eseguire i seguenti passaggi:
 - (a) Aprire il progetto **A-Clus_Base**
 - (b) Recarsi nella sezione **Referenced Libraries**, in asso a sinistra
 - (c) Posizionare il cursore sulla cartella e selezionare **Add JAR/Folder Classpath**, che apparirà sulla destra



3. Avviare **MultiServer.java**
4. Controllare che il server sia in esecuzione correttamente
5. Aprire il progetto **MainTest.java** presente nella cartella **A-CLus_Base/Client**

3.1 Interfaccia iniziale

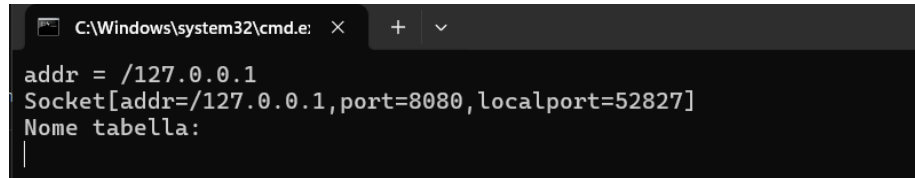
L'interfaccia iniziale di A-CLus è composta da tre sezioni principali:



```
C:\Windows\system32\cmd.e: x + v
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52827]
Nome tabella:
```

Figura 8: Interfaccia iniziale di A-CLus

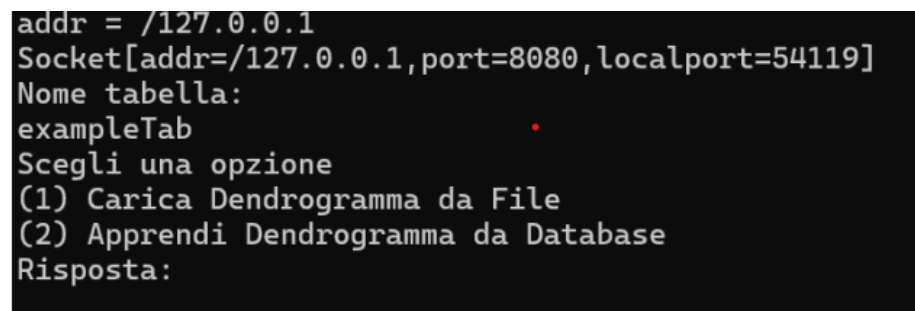
L'interfaccia iniziale di A-CLus è composta da tre sezioni principali:

A screenshot of a Windows command prompt window. The title bar shows 'C:\Windows\system32\cmd.exe'. The command prompt contains the following text: 'addr = /127.0.0.1', 'Socket[addr=/127.0.0.1,port=8080,localport=52827]', and 'Nome tabella:' followed by a cursor on a new line.

```
C:\Windows\system32\cmd.exe
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52827]
Nome tabella:
|
```

Figura 9: Interfaccia iniziale di A-CLus

3.2 Selezione delle operazioni disponibili

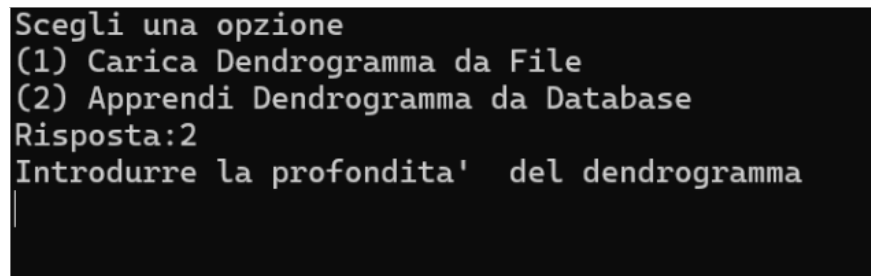
A screenshot of the A-CLus interface in a command prompt. It shows the same initial setup as Figure 9, but with additional text: 'exampleTab', 'Scegli una opzione', a list of two options '(1) Carica Dendrogramma da File' and '(2) Apprendi Dendrogramma da Database', and 'Risposta:' followed by a cursor.

```
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=54119]
Nome tabella:
exampleTab
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:
|
```

Dopo l'inserimento della tabella, verrà proposto un menu con due alternative:

1. **Carica Dendrogramma da File:** permette di importare un dendrogramma precedentemente memorizzato
2. **Apprendi Dendrogramma da Database:** consente di generare un nuovo dendrogramma analizzando i dati presenti nella tabella selezionata

3.3 Percorso operativo - Opzione 2 (Generazione nuovo dendrogramma)

A screenshot of the A-CLus interface in a command prompt. It shows the menu from the previous screenshot, but with 'Risposta:2' entered and a new prompt 'Introdurre la profondita' del dendrogramma' followed by a cursor.

```
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:2
Introdurre la profondita' del dendrogramma
|
```

Selezionando l'opzione 2, il sistema richiederà l'inserimento del parametro di profondità desiderato per il dendrogramma.

3.4 Inserimento Profondità

```
Introdurre la profondita' del dendrogramma
3
Distanza: single-link (1), average-link (2):
```

Successivamente, l'applicazione proporrà due metodologie di calcolo alternative:

1. **Single-link:** identifica la distanza minima tra cluster, connettendo gli elementi più prossimi tra i gruppi
2. **Average-link:** determina la distanza utilizzando la media tra tutti gli elementi dei cluster considerati

3.4.1 Elaborazione con Single-link

```
Distanza: single-link (1), average-link (2):
1
level0:
cluster0:<[1.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>
cluster4:<[2.0,2.0,0.0]>

level1:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>

level2:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]><[1.0,3.0,4.0]>

Inserire il nome dell'archivio (comprensivo di estensione):
```

Selezionando l'opzione Single-link (1), verrà elaborato e visualizzato il dendrogramma risultante. L'applicativo richiederà poi di specificare il nome del file per l'archiviazione del dendrogramma, completo di estensione.

3.5 Inserimento nome file

```
Inserire il nome dell'archivio (comprensivo di estensione):
Example1.dat
Dendrogramma salvato correttamente.
Premere un tasto per continuare . . . |
```

Indicare il nome desiderato (esempio: `Example1.dat`). Il dendrogramma verrà salvato nella cartella "`saved`" all'interno della directory "`Jar + Bat`" e l'applicazione terminerà l'esecuzione.

3.5.1 Elaborazione con Average-link

```
Distanza: single-link (1), average-link (2):
2
level0:
cluster0:<[1.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>
cluster4:<[2.0,2.0,0.0]>

level1:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>

level2:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]><[1.0,3.0,4.0]>

Inserire il nome dell'archivio (comprensivo di estensione):
```

Selezionando l'opzione Average-link (2), il sistema elaborerà il dendrogramma utilizzando il criterio della distanza media e lo visualizzerà a schermo. Successivamente, verrà richiesto di specificare il nome del file per la memorizzazione del dendrogramma, completo di estensione.

3.6 Percorso operativo - Opzione 1 (Caricamento dendrogramma esistente)

```
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:1
Inserire il nome dell'archivio (comprensivo di estensione):
```

Selezionando l'opzione 1 per importare un dendrogramma preesistente, sarà necessario indicare il nome completo del file archivio, comprensivo di estensione. I file precedentemente salvati sono localizzati nella cartella "`saved`" all'interno della directory "`Jar + Bat`".

```
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:1
Inserire il nome dell'archivio (comprensivo di estensione):
Example1.dat
level0:
cluster0:<[1.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>
cluster4:<[2.0,2.0,0.0]>

level1:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]>
cluster3:<[1.0,3.0,4.0]>

level2:
cluster0:<[1.0,2.0,0.0]><[2.0,2.0,0.0]>
cluster1:<[0.0,1.0,-1.0]>
cluster2:<[1.0,3.0,5.0]><[1.0,3.0,4.0]>

Premere un tasto per continuare . . .
```

3.7 Inserimento nome archivio con estensione

Dopo l'inserimento del nome dell'archivio, il sistema visualizzerà il dendrogramma caricato e concluderà l'esecuzione.

4 TEST A-CLUS BASE

Durante la fase di testing, sono stati analizzati diversi scenari critici per validare la robustezza dell'applicazione:

4.1 Gestione di tabelle inesistenti

```
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=54675]
Nome tabella:
example
Errore nel recupero dei dati: Table 'mapdb.example' doesn't exist
Nome tabella:
```

Quando viene specificato un identificativo di tabella non presente nel database, il sistema visualizza un appropriato messaggio diagnostico e offre all'utente la possibilità di inserire una denominazione alternativa.

4.2 Validazione delle selezioni di menù

La schermata di selezione del menù principale consente di scegliere tra due opzioni: "Carica Dendrogramma" o "Esegui Clustering".

```
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:3
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:
```

L'inserimento di valori non conformi alle opzioni disponibili (diversi da 1 o 2) viene intercettato, consentendo all'utente di ripetere la selezione senza interruzioni del flusso operativo.

4.2.1 Controllo sull'esistenza degli archivi

```
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:1
Inserire il nome dell'archivio (comprensivo di estensione):
example.dat
File non trovato: example.dat
Premere un tasto per continuare . . .
```

Nel caso di caricamento di un dendrogramma precedentemente salvato, il sistema verifica l'esistenza dell'archivio specificato. In caso negativo, viene presentata una notifica e l'applicazione termina la sessione.

4.2.2 Validazione parametri di profondità

```
(2) Apprendi Dendrogramma da Database
Risposta:2
Introdurre la profondita' del dendrogramma
0
Distanza: single-link (1), average-link (2):
1
E' stato inserito un valore <=0 per la profondita' del Dendrogramma.
Premere un tasto per continuare . . .
```

L'inserimento di valori non ammissibili per il parametro di profondità (zero o negativi) viene rilevato dal sistema che, pur proseguendo con la richiesta del metodo di calcolo, interromperà l'elaborazione notificando l'anomalia.

```
(2) Apprendi Dendrogramma da Database
Risposta:2
Introdurre la profondita' del dendrogramma
-2
Distanza: single-link (1), average-link (2):
2
E' stato inserito un valore <=0 per la profondita' del Dendrogramma.
Premere un tasto per continuare . . .
```

4.2.3 Controllo selezioni metodologiche

```
Distanza: single-link (1), average-link (2):
3
Distanza: single-link (1), average-link (2):
```

L'inserimento di opzioni non valide per la selezione della metodologia di calcolo delle distanze viene gestito permettendo all'utente di riformulare la propria scelta.

4.2.4 Verifica di compatibilità dimensionale

```
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=54135]
Nome tabella:
exampleTab1
Scegli una opzione
(1) Carica Dendrogramma da File
(2) Apprendi Dendrogramma da Database
Risposta:1
Inserire il nome dell'archivio (comprensivo di estensione):
esempiol11.dat
Il numero di esempi nella tabella scelta e' minore del numero di esempi con cui e' stato salvato l'oggetto h-clus
precedentemente serializzato.
Premere un tasto per continuare . . .
```

Il sistema controlla la congruenza tra la cardinalità degli esempi nella tabella corrente e quella memorizzata nell'oggetto serializzato. In caso di incongruenza, viene visualizzato un messaggio esplicativo e l'esecuzione viene terminata.

4.3 Gestione connettività client-server

```
C:\WINDOWS\system32\cmd. × + v
addr = /127.0.0.1
Connection refused: connect
Premere un tasto per continuare . . . |
```

L'avvio del client in assenza del componente server attivo genera un tentativo di connessione all'indirizzo locale (127.0.0.1) che, non trovando un endpoint disponibile, risulta in un rifiuto della connessione con relativa notifica.

4.3.1 Elaborazione su dataset vuoti

```
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52441]
Nome tabella:
exampleTab1
Errore nel recupero dei dati: La tabella exampleTab1 è vuota.
Nome tabella:
|
```

Quando l'utente tenta di eseguire un'analisi di clustering su una tabella priva di contenuti, il sistema identifica la condizione e fornisce un'appropriata segnalazione.

5 Eseguire A-CLus Esteso

Per avviare la versione estesa del progetto A-CLus, è necessario seguire i seguenti passaggi:

1. Accedere alla directory `H-CLUS_Esteso/Jar + Bat`
2. All'interno della suddetta cartella, eseguire il file `Start Server.bat` con doppio click, mantenendo attiva la finestra del prompt dei comandi
3. Procedere con l'avvio del file `Start Client.bat` tramite doppio click, il quale genererà un'ulteriore finestra di comando
4. Aprire Telegram e cercare "A-Clus" utilizzando il tag "@A_Clus.bot" oppure cliccando sul link diretto