Sommario

[**1. Introduzione** 2](#_Toc186287499)

[**2. Hierarchical Clustering: Introduzione al progetto** 2](#_Toc186287500)

[**2.1** **Estensione sviluppata** 3](#_Toc186287501)

[**2.2** **Progettazione** 3](#_Toc186287502)

[**3. Diagrammi UML** 5](#_Toc186287503)

[**3.1 Server** 5](#_Toc186287504)

[**3.2 Client console** 6](#_Toc186287505)

# [1. Introduzione](#_1._Introduzione)

La tecnica del **Hierarchical Clustering** è una delle tecniche di clustering più utilizzate nel data mining, in grado di creare una rappresentazione gerarchica dei cluster in un set di dati. L’algoritmo utilizzato nel progetto è il clustering gerarchico di tipo agglomerativo in cui ogni punto dati viene considerato come un cluster individuale. A ogni iterazione, vengono combinati i cluster più vicini fino a raggiungere un criterio di arresto.

Il risultato del clustering agglomerativo dipenderà dalle tecniche utilizzate per calcolare la distanza tra i cluster. I metodi impiegati per tale calcolo sono la distanza Single-Link e la distanza Average-Link. Nel metodo del Single-Link la distanza tra due cluster è definita come la distanza minima tra i punti di due cluster diversi, invece nel metodo dell’Average-Link la distanza tra due cluster è definita come la media delle distanze tra tutti i punti di un cluster e tutti i punti dell'altro cluster.

Il risultato del clustering gerarchico è una struttura ad albero, chiamata dendrogramma, che rappresenta le relazioni gerarchiche tra i cluster.

# 2. [Hierarchical Clustering: Introduzione al progetto](#_Hierarchical_Clustering:_Introduzio)

L’algoritmo di clustering agglomerativo è stato implementato in un software Java progettato per eseguire la clusterizzazione di esempi di dati estratti da una tabella di un database. Il software è composto da due applicativi: un client e un server. In particolare, il client funge da interfaccia per l'utente, permettendo di interagire con il server tramite terminale. Inoltre, il client permette di avviare operazioni di clustering e gestire il caricamento o la lettura dei file generati. Dunque, offre funzionalità intuitive per selezionare le impostazioni desiderate, come la profondità del dendrogramma da definire, il tipo di distanza tra i cluster da utilizzare e la tabella da cui estrarre i dati. Il server è il motore vero e proprio del software in quanto può eseguire diverse richieste per i diversi client connessi. In particolare, può estrapolare dati da una tabella di un database, eseguire il clustering su quei dati, salvarli e di conseguenza leggerli da un file. Inoltre, lo stato del server è consultabile da terminale.

# [Estensione sviluppata](#_Estensione_sviluppata)

Nonostante la versione base sia completa e funzionante, per rendere il progetto accessibile a più utenti è stata sviluppata, come estensione, un’interfaccia grafica per il client. Questa ha reso il progetto più intuitivo e user-friendly rispetto a un sistema che richiede l'interazione attraverso un terminale. L’aspetto dell’interfaccia è semplice e pulito garantendo all’utente facilità d’uso e maggiore chiarezza e libertà di navigazione tra le varie finestre senza ambiguità e possibilità di cadere in errore.

# [Progettazione](#_Progettazione)

L’interfaccia grafica è stata costruita utilizzando **Vaadin Flow**, la versione moderna di Vaadin, un framework Java per la creazione di applicazioni web con interfacce utente interattive e responsive che genera automaticamente il codice HTML e JavaScript per il browser. Il programma utilizza anche il framework open source **Spring Boot** come base per configurare e avviare il server, integrandosi perfettamente con Vaadin. Per personalizzare l’aspetto grafico dell’interfaccia sono stati usati file .css per definire e applicare stili ai singoli elementi, e file .svg dal momento che l’interfaccia include due icone.

Dopo aver definito con chiarezza i requisiti e valutato possibili fonti di ispirazione per la struttura dell'interfaccia grafica da sviluppare, si è proceduto alla realizzazione degli elementi costitutivi del progetto. Il lavoro è stato organizzato in modo collaborativo, con una suddivisione delle attività tra i due membri del gruppo: ciascuno si è concentrato su aspetti differenti dell’interfaccia, lavorando rispettivamente sul funzionamento e sull'aspetto grafico. Le parti sviluppate individualmente sono state integrate iterativamente, consentendo un continuo allineamento e un miglioramento progressivo dell’intero sistema.

Il progetto dell'interfaccia ha coinvolto la realizzazione di tre schermate principali. La prima schermata raffigura la home del progetto in cui avviene l'avvio del programma e la connessione al server. Include una descrizione del progetto e due opzioni per l'utente. A seconda della scelta, la pagina può indirizzarlo verso l'individuazione dei cluster da un database oppure verso la lettura di dati da un file.

La realizzazione del progetto è visibile nella guida utente di questo documento.

# [3. Diagrammi UML](#_3._Diagrammi_UML)

In questa sezione del documento vengono mostrati i diagrammi UML delle classi e dei package. Tutti i diagrammi sono stati realizzati utilizzando il software StarUML. Abbiamo tratto ispirazione dal libro Java Modeling Color with UML e quindi deciso di utilizzare i colori per identificare le classi di uno stesso package.

# [3.1 S](#_3._Diagrammi_UML)erver

**Diagramma dei package**

Immagine che contiene testo, diagramma, linea, schermata

Descrizione generata automaticamente

Nello specifico:

Immagine che contiene testo, schermata, Carattere, Rettangolo

Descrizione generata automaticamente

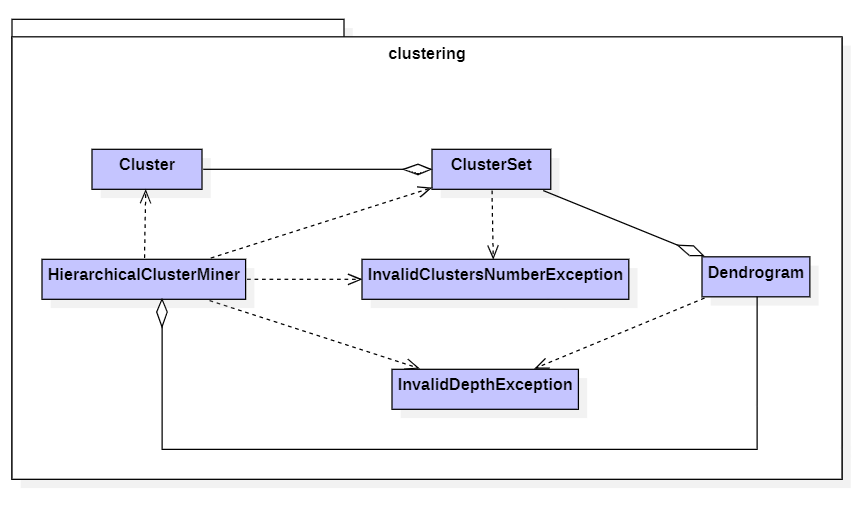


Immagine package distance

Immagine package data

Immagine package database

**Diagramma delle classi**

Immagine diagramma delle classi

# [3.2](#_3._Diagrammi_UML) Client console

**Diagramma dei package**

Immagine + immagini package specifici

**Diagramma delle classi**

Immagini

Manuale Utente

Requisiti

Per poter eseguire il software H-CLUS non sono necessarie procedure complesse. Per il corretto funzionamento è necessario seguire i seguenti passaggi:

* Scaricare e installare una versione aggiornata di **Java Development Kit (JDK)**. Assicurarsi di aver installato il JDK versione 22 o successiva.
* Scaricare e installare **MySQL**, assicurarsi di installare una versione 5.7 o successiva di MySQL Community Server.
* Avviare e verificare che il servizio MySQL sia in esecuzione.
* Eseguire lo script “**setup.sql**” per configurare il database con MySQL.

Versione Console

Istruzioni per l’esecuzione

Server

Prima di utilizzare il software, assicurarsi di aver rispettato i requisiti del punto precedente. È necessario che venga avviato il server prima di avviare il client. In caso contrario verrà visualizzato un messaggio di errore, interrompendo l’esecuzione del programma client.

Nella cartella principale del progetto individuare il file ***run\_server.bat*** e prima di avviarlo, modificare il file sostituendo il valore del campo *port* col numero di porta sul quale avviare il server. Lasciare invariato il campo se si preferisce utilizzare il valore di default.

Per poter avviare il Server in maniera automatica è possibile utilizzare il file ***run\_server.bat***, infatti basta semplicemente individuare il file nella cartella del progetto e fare doppio click per eseguirlo. In alternativa si può aprire un terminale o prompt dei comandi, navigare nella directory dove si trova il file ed eseguire il comando: *run\_server.bat*

È possibile avere aperto un unico Server alla volta, infatti il sistema impedirà di aprirne più di uno contemporaneamente.

Client

Nella cartella principale del progetto individuare il file ***run\_client.bat*** e prima di avviarlo, modificare il file sostituendo i valori dei campi *address* e *port* con l’indirizzo ip e il numero di porta del server avviato in precedenza. Lasciare invariati i campi se si preferisce utilizzare i valori di default.

Una volta completata la configurazione è possibile avviare il Client in maniera automatica utilizzando il file ***run\_client.bat***, infatti basta semplicemente individuare il file nella cartella del progetto e fare doppio click per eseguirlo. In alternativa si può aprire un terminale o prompt dei comandi, navigare nella directory dove si trova il file ed eseguire il comando: *run\_client.bat.*

Guida utente

Aprendo run\_server.bat, avviene l’avvio del server. Visualizzando il messaggio in figura, il server è pronto a ricevere richieste dai client.

SCREEN AVVIO SERVER

Aprendo run\_client.bat, avviene l’avvio del client. Vengono visualizzate le informazioni di connessione e nome tabella dove bisogna inserire il nome della tabella su cui si vuole lavorare, in questo caso *exampleTab*.

SCREEN AVVIO CLIENT

L’utente può quindi scegliere tra due opzioni: caricare il dendrogramma da un file oppure apprendere da una tabella da database. Scegliendo l’operazione 2 dal menu