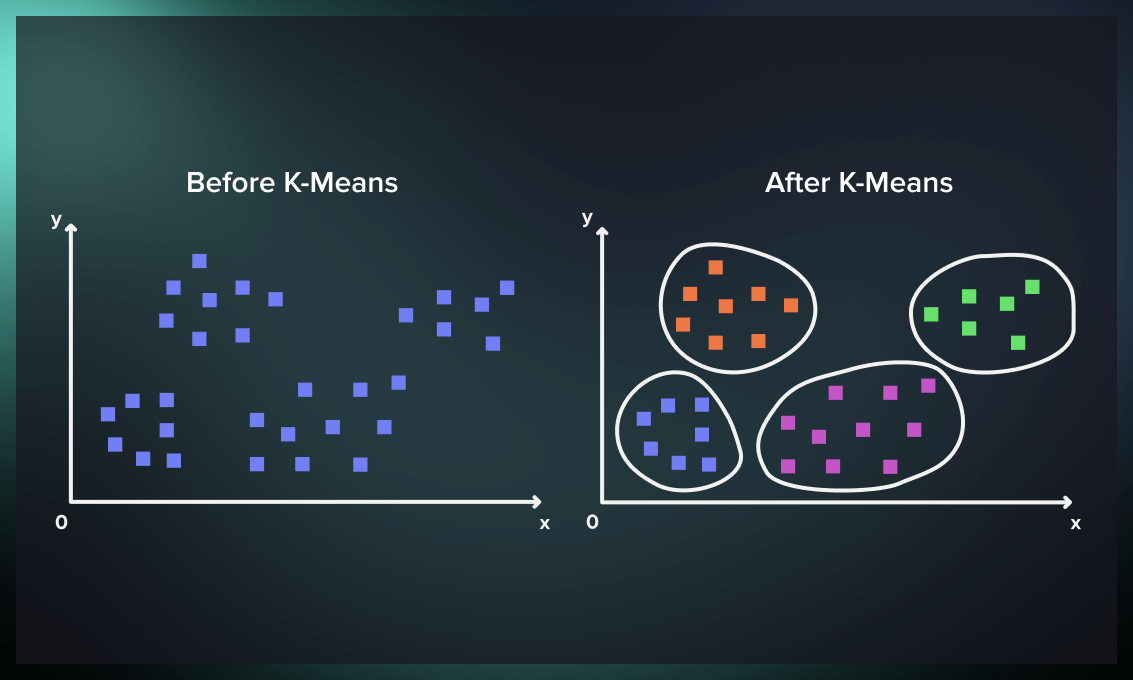
Progetto Map 2022-23:

K-MEANS

Clustering



1. **Introduzione**

Il caso di studio analizzato nell’anno 2022-2023 ha visto come soggetto il clustering cioè, dati una collezione D di transizioni dove, ogni transazione è un vettore di coppie attributo-valore (item), e un intero k, lo scopo è partizionare D in k insiemi di transazioni , tale che è un segmento (selezione) omogenea di D, e .

Per eseguire il clustering bisogna utilizzare il K-means: un algoritmo di analisi dei gruppi partizionale che permette di suddividere un insieme di oggetti in k gruppi sulla base dei loro attributi.

Per eseguire il K-means bisogna:

1. Selezionare il numero di cluster k che vogliamo identificare
2. Selezionare k punti casuali che fungeranno da centroidi dei cluster che si creeranno
3. Creare k cluster misurando la distanza di ogni punto da ciascun dei k centroidi, quindi assegnare i punti al cluster più vicino.
4. Calcolare il nuovo centroide di ciascun cluster utilizzano i cluster ottenuti
5. Valutare la qualità di ciascun cluster trovando la variazione all’interno di tutti i cluster, definendo i cluster in modo che essa sia ridotta al minimo
6. Una volta memorizzati i cluster precedenti e la variazione, ripetere i passi 3-5 ma utilizzando i centroidi calcolati in precedenza per creare k nuovi cluster, ricalcolare il centro dei nuovi cluster e calcolare la somma delle variazioni all’interno di tutti i cluster.

Il valore di k è compreso tra 1 (la variazione diventa pari a 0) e il numero totale degli item facendo diventare ogni item un cluster.

1. **Da chi è stato realizzato il progetto**

* **Boffolo Alessandro Aldo** (mat.:735963)

E-mail: [a.boffolo@studenti.uniba.it](mailto:a.boffolo@studenti.uniba.it)

* **Canonico Luca** (mat.:708794)

E-mail: [l.canonico@studenti.uniba.it](mailto:l.canonico@studenti.uniba.it)

* **Fornaro Emanuele** (mat.:738721)

E-mail: [e.fornaro4@studenti.uniba.it](mailto:e.fornaro4@studenti.uniba.it)

1. Come è stato realizzato il progetto

La struttura del progetto si basa sul linguaggio Java, inoltre è stato utilizzato MySQL Workbench 8.0 CE per poter leggere i data target dalla base di dati.

L’interfaccia utente è stata creata con JavaFX e semplifica l’utilizzo del programma, rendendolo più chiaro, grazie all’inserimento di elementi grafici e di finestre.

1. Funzionalità implementate

**Versione base**

Nella versione base del programma è possibile leggere o scoprire i vari cluster da una base di dati, attraverso una scelta che viene effettuata sottoforma di menu testuale.

Se si sceglie di scoprire i cluster, si dovrà inserire il nome della tabella all’interno della base di dati da cui estrapolare i dati e, in seguito, inserire il numero di cluster da visualizzare.

Una volta visualizzati è anche possibile salvare i dati in un file, dando prima il nome, che potranno essere caricati in seguito con l’opzione di lettura.

Con l’opzione lettura è possibile visualizzare un file contenente i cluster scoperti in precedenza, inserendo il nome di tale file.

È stata implementata, inoltre, la gestione degli errori, attraverso la quale si cerca di spiegare nel modo più chiaro possibile all’utente eventuali problemi riscontrati durante la comunicazione tra client e server e/o trasmissioni con il database fornendo una soluzione.

Essendo inoltre il server multithreading, è consentita una comunicazione tra un server e più client, in tempo reale e con connessioni multiple, grazie all’estensione della classe Thread di Java.

**Versione estesa**

Nella versione estesa il programma offre la possibilità di scoprire o leggere i vari cluster da una base di dati attraverso l’uso di una interfaccia grafica.

All’avvio viene richiesto di inserire l’Indirizzo IP (127.0.0.1) e il Numero di Porta (8080), dopo di che si passa a una nuova schermata dove si potrà scegliere se scoprire o leggere i cluster.

Nel caso si vogliano scoprire i cluster si prosegue inserendo prima il nome della tabella all’interno della base di dati e inseguito il numero di cluster da visualizzare.

Nel caso si scelga invece l’opzione di lettura, bisognerà inserire il nome del file salvato contente i cluster già scoperti ed eseguirlo per poterlo visualizzare.

1. Guida Utente

**PREREQUISITI**

**Java** – Per l’esecuzione del programma (client o server, versione base o estesa) è necessario che sul computer sia stato installato Java e che la versione dell’ambiente Java (JRE) sia uguale o superiore alla 1.8.

Link di download Java: <https://www.java.com/it/download/>

**VERSIONE BASE**

Il K-means si compone di due programmi: un programma client e un programma server. Per poter permettere la comunicazione bisogna avviare in primis l’applicativo server su un qualsiasi host.

**Preconfigurazione**

1. Avvio MultiServer (**Server**)

Per avviare "MultiServer" aggiungere alla propria configurazione di Run un numero di porta valida su cui ospitare il servizio Server (valori al di fuori del range 1-1024 e con un massimo di 65535).

Nella configurazione fornita il numero di porta settato sarà 8080, se la si vuole cambiare seguire i seguenti passi: "Edit configurations...", nella sezione "Program arguments" inserire un valore valido per esempio **8080** -> premere su “Apply” per applicare le modifiche.

In caso di "Errore di connessione al database: com.mysql.cj.jdbc.Driver" assicurarsi di aver collegato il Driver di MySQL, eseguire i seguenti passi:

- File -> Project Structure -> Modules -> KmeansServer -> Dependencies -> + -> JARs or Directories… -> navigate\to\the\driver\path -> premere su "mysql-connector-java-8.0.17.jar" ->Apply

Infine assicurarsi di avere eseguito lo Script MySQL “createDB” sulla macchina che ospita il Server, in modo da caricare all’interno del DataBase la tabella.

1. Avvio MainTest (**Client**)

Per avviare "MainTest" aggiungere alla propria configurazione di Run un IP address e un numero di porta su cui il Server è in ascolto. La configurazione fornita presenta come indirizzo IP localhost “127.0.0.1” e il numero di porta “8080”.

Per cambiare la configurazione seguire i seguenti passi: Edit configurations...", nella sezione "Program arguments" inserire l’indirizzo IP e il numero di porta su cui il Server fornisce il servizio es. -> **127.0.0.1 8080** -> premere “Apply”

1. Avvio Estensione (**ClientExtension**)

In caso di errore” JavaFX runtime components are missing…” assicurarsi di avere javafx-sdk collegato (incluso nel file del progetto). Seguire i seguenti passi:

1. Create a JavaFX library in Idea

- File -> Project Structure -> Libraries -> + -> Point to the lib folder of the JavaFX SDK -> Apply

2. Add VM options for the module path

- Run -> Edit Configurations -> Modify options -> Add VM options -> --module-path out\artifacts\Extension\_jar\javafx-sdk-21.0.1\lib --add-modules javafx.controls,javafx.fxml,javafx.graphics,javafx.base -> Apply

**Avvio dell’applicativo (versione Base)**

Una volta configurato, se si vuole aprire il programma, bisognerà aprire le seguenti cartelle: out -> artifacts -> KmeansServer\_jar

Ed eseguire prima il file **KmeansServer.bat,** che mostrerà questa schermata:

Immagine che contiene testo, schermata, software, Software multimediale

Descrizione generata automaticamente

In seguito, per avviare la versione base, bisognerà avviare il file **KmeansClient.bat**, attraverso il percorso: out -> artifacts -> KmeansClient\_jar.

Nel menu che comparirà si potrà scegliere se leggere o scoprire i vari cluster da una base di dati.

Immagine che contiene testo, schermata, schermo, software

Descrizione generata automaticamente

**Carica Dati**

Se si vogliono scoprire i cluster bisognerà digitare **2**, si dovrà poi inserire il nome della tabella all’interno della base di dati da cui estrapolare i dati e, in seguito, inserire il numero di cluster da visualizzare.

Immagine che contiene testo, schermata, schermo, software

Descrizione generata automaticamente

Una volta visualizzati è anche possibile salvare i dati in un file, attribuendogli un nome (come nell’esempio qui sotto dove il file è stato chiamato **uno**).

Immagine che contiene testo, schermata, Carattere, numero

Descrizione generata automaticamente

**Carica Cluster da File**

Invece, digitando **1** nel menu iniziale, è possibile visualizzare un file contenente i cluster scoperti in precedenza, inserendo il nome di tale file (es. **uno**).

Immagine che contiene testo, schermata, Carattere, software

Descrizione generata automaticamente

**Avvio dell’applicativo (versione Estesa)**

La versione estesa presenta le stesse funzionalità del programma base, ma con l’implementazione di un’interfaccia grafica, il cui scopo è quello di rendere l’applicativo user-friendly. Se la si vuole avviare bisognerà eseguire il file **KmeansServer.bat** attraverso il percorso: out -> artifacts -> KmeansServer\_jar

Ed in seguito eseguire il file **Extension.bat** attraverso il percorso: out -> artifacts -> Extension\_jar che mostrerà la seguente schermata:

Immagine che contiene testo, schermata, software, Pagina Web

Descrizione generata automaticamente

Qui viene richiesto di inserire l’Indirizzo IP (es. 127.0.0.1) e il Numero di Porta (es. 8080). Una volta inseriti bisognerà premere sul pulsante *Connetti*.

Dopodiché, se l’indirizzo IP e il Numero di Porta sono validi, visualizzeremo una nuova schermata (Schermata principale) dove si potrà scegliere se scoprire o leggere i cluster.

Immagine che contiene testo, schermata, software, schermo

Descrizione generata automaticamente

Premendo il pulsante *Scoperta* verrà visualizzata la seguente schermata dove sarà richiesto di inserire il nome della tabella all’interno della base di dati dalla quale scoprire i cluster.

Immagine che contiene testo, schermata, software, Sistema operativo

Descrizione generata automaticamente

Premendo il pulsante *Inserisci* si passa a una nuova schermata dove verrà richiesto il numero di cluster da visualizzare.

Immagine che contiene testo, schermata, software, Sistema operativo

Descrizione generata automaticamente

Premendo su pulsante *Esegui* verrà aperta una nuova finestra contenente i cluster scoperti, che potranno essere salvati in un file, inserendo il nome del file, come nell’immagine sottostante, e premendo il pulsante *Salva.*

Immagine che contiene testo, elettronica, schermata, schermo

Descrizione generata automaticamente

Infine, basterà chiudere questa finestra per poter tornare a quella precedente, dove si potrà inserire un nuovo valore di cluster da visualizzare all’interno della stessa tabella, oppure si potrà tornare alla schermata principale per poter scegliere se leggere o scoprire i cluster premendo sul pulsante a forma di casetta in basso a destra.

Premendo il pulsante *Lettura* nella Schermata Principalesarà possibile visualizzare un file contenente i cluster scoperti in precedenza, inserendo il nome di tale file e premendo il pulsante *Esegui* come nella schermata seguente.

Immagine che contiene schermata, testo, schermo, software

Descrizione generata automaticamente

Nella schermata successiva viene visualizzato il file.

Immagine che contiene testo, elettronica, schermata, schermo

Descrizione generata automaticamente

Premendo sul pulsante a forma di casetta si tornerà alla Schermata Principale.

**Tipologie di Errori**

*Errore server non attivo*

Aprendo prima il file Client del file Server verrà visualizzato il seguente messaggio

Immagine che contiene testo, schermata, software, schermo

Descrizione generata automaticamente

*Errore nome della tabella di database*

Inserendo nel menu un nome di una tabella di database non presente verrà visualizzato il seguente messaggio

Immagine che contiene testo, elettronica, schermata, schermo

Descrizione generata automaticamente

*Errore numero cluster da scoprire*

Inserendo un valore numerico di cluster da voler scoprire maggiore o minore di quello consentito verrà visualizzato il seguente messaggio

Immagine che contiene testo, elettronica, schermata, schermo

Descrizione generata automaticamente

*Errore numero di cluster da scoprire inserendo una lettera*

Inserendo un valore non numerico di cluster da voler scoprire verrà visualizzato il seguente messaggio

Immagine che contiene testo, elettronica, schermata, schermo

Descrizione generata automaticamente

*Errore nome file in cui sono serializzati i cluster da recuperare*

Inserendo il nome di un file non presente tra quelli salvati dalla quale visualizzare i cluster scoperti verrà visualizzato il seguente messaggio

Immagine che contiene testo, elettronica, schermata, schermo

Descrizione generata automaticamente

*Errore Indirizzo IP non valido (Versione Estesa)*

Inserendo un Indirizzo IP non valido verrà visualizzato il seguente messaggio

Immagine che contiene testo, schermata, Carattere, software

Descrizione generata automaticamenteImmagine che contiene testo, schermata, software, schermo

Descrizione generata automaticamente

*Errore Numero di Porta (versione Estesa)*

Inserendo un Numero di Porta non valido verrà visualizzato il seguente messaggio

Immagine che contiene testo, schermata, Carattere, software

Descrizione generata automaticamenteImmagine che contiene testo, schermata, software, schermo

Descrizione generata automaticamente

*Errore nome della tabella di database (versione Estesa)*

Inserendo nel menu un nome di una tabella di database non presente verrà visualizzato il seguente messaggio

Immagine che contiene testo, schermata, schermo, software

Descrizione generata automaticamenteImmagine che contiene testo, schermata, software, schermo

Descrizione generata automaticamente

*Errore numero cluster da scoprire (versione Estesa)*

Inserendo un valore numerico di cluster da voler scoprire maggiore o minore di quello consentito verrà visualizzato il seguente messaggio

Immagine che contiene testo, schermata, Carattere, software

Descrizione generata automaticamenteImmagine che contiene testo, schermata, software, schermo

Descrizione generata automaticamente

*Errore numero di cluster da scoprire inserendo una lettera (versione Estesa)*

Inserendo un valore non numerico di cluster da voler scoprire verrà visualizzato il seguente messaggio

Immagine che contiene testo, schermata, software, Carattere

Descrizione generata automaticamenteImmagine che contiene testo, schermata, software, schermo

Descrizione generata automaticamente

*Errore nome file in cui sono serializzati i cluster da recuperare (versione Estesa)*

Inserendo il nome di un file non presente tra quelli salvati dalla quale visualizzare i cluster scoperti verrà visualizzato il seguente messaggio

Immagine che contiene testo, schermata, schermo, software

Descrizione generata automaticamenteImmagine che contiene schermata, testo, schermo, software

Descrizione generata automaticamente