Buongiorno, vi diamo il benvenuto alla revisione di accettazione,

\*Slide\*Noi siamo il gruppo Codebusters ed andremo a esporre il nostro prodotto.

**Elevator pitch:**

**Problema:**

\*Slide\*Iniziamo analizzando il problema: oggigiorno molte aziende, dai colossi come Google, Amazon, Twitter, alle imprese presenti nel nostro territorio, devono effettuare analisi su dei dati che spesso sono eterogenei. Dunque, per ricavare informazioni utili hanno bisogno di particolari strumenti che permettano di elaborarli e visualizzarli in modo appropriato.

\*Slide\*Per comprendere meglio il dominio del problema facciamo un esempio di dataset multidimensionale. Come si vede nella slide per ogni persona si hanno dei dati diversi relativi al sesso, livello di istruzione, figli, età, se è un fumatore e quante sigarette fuma. Tutte queste informazioni in genere vengono definite dimensioni del dataset. Questo dataset possiede quindi dei dati eterogenei, difficili da visualizzare e comprendere appieno a primo impatto, in particolare nel caso in cui fosse composto da molte righe ovvero molti campionamenti.

\*Slide\*Questa problematica, ad esempio, può presentarsi nel caso in cui un’azienda, che vuole comprendere quale sia il target a cui è rivolto un loro prodotto, debba analizzare una vasta quantità di dati multidimensionali, potenzialmente forniti da questionari e analisi di mercato. Un altro scenario è quello di un’azienda che possiede un software per la gestione del personale e deve verificare se tutti i dati inseriti manualmente, e quindi soggetti a frequenti errori, sono corretti o se ce ne sono di anomali.

Questi due esempi hanno in comune il fatto che i dati da analizzare sono di diversa natura e devono essere confrontati per poter raggiungere un risultato significativo da cui ricavare le informazioni necessarie.

**Prodotti concorrenti:**

\*Slide\*In commercio ci sono alcuni strumenti come Excel, Orange Canvas, Ggobi o rapidminer che permettono l’analisi di dati a molte dimensioni con l’aiuto di grafici.

Excel non è adatto a grandi moli di dati, perché è un foglio di calcolo e il suo scopo non è la visualizzazione a livello multidimensionale.

Orange Canvas, invece, è un software macchinoso che richiede procedimenti inutilmente prolissi.

Gli altri software citati, nonostante svolgano il loro compito, a volte risultano poco intuitivi per chi non è avvezzo ad usare questi strumenti, utilizzando terminologie o funzionalità di difficile interpretazione, rendendo il procedimento di analisi dei dati un lavoro lungo, stressante e complesso, ma evitabile. Come è evitabile?

**Soluzione:**

\*Slide\*La soluzione che noi proponiamo è HD Viz: una web app che riesce a visualizzare dati a molte dimensioni in cinque diversi tipi di grafico.

\*Slide\*Un tipico utilizzo del programma prevede, inizialmente, il caricamento dei dati tramite file CSV o con interrogazioni ad un database.

A questo punto i dati possono essere studiati: HD Viz, oltre alla visualizzazione dei grafici, fornisce anche algoritmi per la riduzione delle dimensioni e funzioni per il calcolo della distanza, strumenti indispensabili per dataset multidimensionali.

In orange canvas, per esempio, il problema che si presenta dopo aver caricato i dati è comprendere come poter visualizzare un grafico: con HdViz basta un click ed è possibile scegliere una delle diverse visualizzazioni di dati offerte.

\*Slide\*L’utente ha anche la possibilità di interrompere una sessione di lavoro e ripristinarla in un secondo momento tramite la generazione di un file in formato JSON che conterrà i dati e le preferenze di visualizzazione. Questo permette di mantenere salvati dei grafici, con le loro rispettive preferenze di visualizzazione, potendo così anche condividere con altri utenti il proprio lavoro e i risultati ottenuti.

\*Slide\*Le funzionalità fornite da HD Viz sono semplici e intuitive e puoi, volendo, visualizzare dati relativi al numero di utenti di Chrome ed Internet Explorer (scoprendo che c’è qualcuno che lo usa ancora) oppure sui voti nel primo appello di Ingegneria del software.

\*Slide\*E visualizzare il grado di felicità e soddisfazione che proviamo nell’essere riusciti a completare il prodotto? Questo è impossibile.

\*Slide\*“Ci sono cose che non si possono rappresentare, per tutto il resto c'è HD Viz”.

**Demo**

***Giacomo: introduzione***

Andiamo ora quindi a dare una dimostrazione del nostro prodotto.

Questa è la schermata in cui l’utente si trova all’avvio dell’applicazione, abbiamo optato per uno stile semplice e user-friendly, l’applicazione infatti è composta da un’unica schermata che si aggiorna in base a ciò che l’utente desidera fare. In alto a destra spicca un pulsante che permette di visualizzare una guida che spiega brevemente tutte le funzionalità presenti nell’applicazione in cui sono presenti anche i link a dei video-tutorial presenti su YT, ed un altro pulsante per caricare un dataset di prova, nel caso in cui un utente volesse solamente testare la nostra applicazione. A sinistra invece si può vedere il menù, in cui sono presenti tutti i pulsanti per le diverse funzionalità rese disponibili.

Come si può vedere all’inizio un utente ha a disposizione solamente la possibilità di caricare i dati da csv o da db oppure di ripristinare una sessione salvata in qualche utilizzo precedente. Tutte le altre funzionalità sono disattivate e si può vedere passandoci sopra con il mouse il motivo per cui lo sono. Inizialmente tutte queste sono disattivate perchè è necessario aver caricato un dataset nell’applicazione prima di fare qualsiasi altra operazione.

Per provare l’applicazione andiamo quindi a caricare un dataset da csv e in questo caso selezioniamo tutte le dimensioni, eventualmente sarebbe possibile evitare di caricare qualcuna di queste andando semplicemente a togliere la spunta dalla checkbox relativa alla dimensione da non caricare.

***Michele: riduzione dimensionale + plma, hm, spm***

Una volta caricati i dati saranno disponibili tutti i grafici che permettono di visualizzare i dati così come sono o che hanno subito una riduzione dimensionale tramite algoritmo. Partirei proprio da questo tipo di riduzione per poi passare ai grafici interessati per vedere qualche esempio.

***(menu)***

Dal menù quindi clicchiamo l’apposito pulsante...

***(aprire il modale)***

L’utente andrà a scegliere innanzitutto se normalizzare i dati premendo sulla checkbox dedicata, le dimensioni che saranno interessate da questo processo di riduzione, il nome da associare alle nuove dimensioni e poi che algoritmo utilizzare.

Ce ne sono ben 6 a disposizione: fastmap, pca, lle, isomap, t-sne, u-map e come si può notare la maggior parte sono di tipo non lineare in modo che l’utente non perda informazioni rilevanti dal suo dataset. In base alla scelta l’utente ha piena libertà di personalizzazione dei parametri specifici di ciascun algoritmo, o in alternativa tutti saranno impostati a dei valori standard di default.

Basterà poi premere il pulsante “Esegui riduzione” e alla chiusura del modale potrà visualizzare un alert di successo dell’operazione.

Partendo dallo SPM all’utente basterà selezionare le dimensioni che intende visualizzare per veder apparire il grafico, che cambierà in modo dinamico. Per un’analisi ideale potrà selezionarne al massimo 5 contemporaneamente, scegliere che dimensione associare al colore dei punti e, passando con il cursore al di sopra del grafico, visualizzare il valore delle coordinate in alto a destra e queste linee rosse che aiutano la lettura dei valori negli assi ed eventualmente comprendere meglio le relazioni dei dati tra i singoli scatter plot.

Passando all’HM, anche qui potrà selezionare che dimensione associare all’asse X e Y e alla scala di colori di ciascun punto, più o meno intenso in base al suo valore che potrà essere visualizzato passando al di sopra di ciascun rettangolino.

Infine passando al PLMA l’utente andrà a scegliere quali dimensioni inserire nel grafico e nuovamente quale associare al colore dei punti, In questo caso, per studiare meglio il dataset, potrà selezionare con il cursore gli assi e muoverli nello spazio tridimensionale.

**Safy**

[aprire il modale DB]

Andiamo ora a caricare un dataset diverso utilizzando il caricamento tramite database.

L’utente può scegliere il dataset da caricare tra quelli presenti nel database o, in alternativa, può selezionare le dimensioni che desidera caricare. Eventualmente è possibile porre una condizione per la clausola WHERE specificandone la colonna, il segno e il valore. In seguito è possibile inviare la query al server che si occuperà di interrogare il database e di restituire i dati trovati. A conferma dell’invio della query viene quindi visualizzato un messaggio in basso a destra con il numero di dati trovati. A questo punto se l’utente desidera caricare effettivamente i dati all’interno dell’applicazione preme il pulsante conferma, altrimenti può eseguire altre query fintantoché non raggiunge la condizione desiderata. In ogni caso, sia di caricamento riuscito con successo, sia nel caso di fallimento, viene visualizzata un alert in con l’esito dell’operazione.

***AleRago: calcolo della distanza + am, ff***

**(mostrare tooltip)**

Per quanto riguarda le altre visualizzazioni, queste dipendono dal concetto di distanza e non sono quindi utilizzabili senza aver prima calcolato una matrice delle distanze sui dati caricati.

Per fare ciò l’utente può cliccare sulla voce “Calcola distanza” nel menu, la quale..

**(mostrare modale del calcolo della distanza)**

… apre una finestra in cui può:

* normalizzare i dati prima di eseguire effettivamente il calcolo
* selezionare le dimensioni da utilizzare (2 o più)
* scegliere la funzione di calcolo della distanza da utilizzare tra le 4 che forniamo: Euclidea, Canberra, Chebyshev e Manhattan (*spiegare differenza in caso viene chiesta*)
* scegliere infine il nome da dare alla matrice delle distanze da creare, oppure mantenere quello di default

Premendo quindi “Esegui riduzione” verrà memorizzata nel sistema la corrispondente matrice delle distanze e anche in questo caso l’esito dell’operazione verrà segnalato tramite un alert in basso a destra.

**(mostrare Adjacency Matrix)**

A questo punto si sono sbloccati anche le altre due visualizzazione e vedendo per primo l’Adjacency Matrix, attraverso la box delle preferenze a destra, l’utente può selezionare:

* la matrice delle distanze da utilizzare (l’utente infatti ha la possibilità di crearne più di una se ne ha la necessità)
* poi può selezionare la dimensione da utilizzare per l’ordinamento dei punti per visualizzare (con colori differenti, i cluster presenti nei dati in base alla dimensione scelta)
* può decidere quale dimensione associare alle etichette degli assi
* e la distanza massima e minima tra i punti che vuole visualizzare nel grafico, così da poter escludere dalla visualizzazione punti non rilevanti.

(L’Adjency Matrix trasforma la distanza tra i punti in colori più o meno intensi, facendo quindi capire quali oggetti sono più vicini e quali più lontani.)

**(mostrare Force Field)**

L’utente può anche visualizzare il Force Field. Attraverso la box delle preferenze può selezionare:

* anche in questo caso la matrice delle distanze da utilizzare
* la dimensione da utilizzare per la colorazione dei punti
* e infine, come per l’AM, la distanza massima e minima tra i punti che vuole visualizzare

L’utente può interagire direttamente con il grafico selezionando un punto con il cursore e trascinandolo, osservando quindi le forze di attrazione e repulsione che legano i tutti i punti.

Inoltre, per tutte le visualizzazioni, le box delle preferenze hanno sopra di esse un bottone il quale permette di mostrare o nascondere tale box e centralizzare il grafico sullo schermo, per concentrarsi solamente sull’analisi di esso.

**Giacomo: conclusione**

A questo punto, nel caso in cui l’utente volesse salvare lo stato attuale dell’applicazione per recuperarlo in seguito può salvare la sessione.

Cliccando sul pulsante salva/carica sessione può decidere che nome dare alla sessione e premere il pulsante esporta per salvare localmente lo stato attuale dell’applicazione. A questo punto può chiudere l’applicazione e riaprendo, cliccando sempre sullo stesso pulsante può caricare il file salvato localmente e vedere che l’applicazione automaticamente si ripristina allo stato in cui era stata lasciata.

L’applicazione è stata sviluppata in modo da essere facilmente estendibile con ulteriori grafici o algoritmi per la riduzione dimensionale, inoltre è stata altamente ottimizzata per funzionare anche con dataset di dimensioni modeste.

Queste sono tutte le funzionalità da noi sviluppate, la quasi totalità del codice è testata e tutti i test di accettazione concordati con il proponente sono stati superati, siamo quindi ampiamente soddisfatti del lavoro svolto.

Torniamo ora alla presentazione per procedere con l’esposizione.

ALESSANDRO PIROLO

Come si può notare dal consuntivo del periodo appena trascorso, sono risultate necessarie più ore totali rispetto a quelle preventivate. L'avvicinarsi al collaudo ha richiesto più lavoro, specialmente per i verificatori che hanno controllato il prodotto software per scovare eventuali malfunzionamenti e bug, e la documentazione per gli errori grammaticali e di sintassi.

(Slide)

Come si può vedere però dalla tabella del consuntivo finale, questo non ha rappresentato un problema, infatti le ore aggiunte a questi incrementi sono bilanciate da quanto risparmiato nei periodi precedenti. Quindi la variazione non ha modificato il totale conteggio delle ore, che è rimasto invariato da quanto pianificato nelle fasi iniziali del progetto e siamo riusciti a concludere le attività con un bilancio leggermente in positivo.

(Slide)

I membri del gruppo hanno svolto lo stesso monte ore di lavoro, come si può notare dalla tabella. I ruoli che hanno subito più modifiche durante gli incrementi sono quelli del verificatore e del responsabile. Il primo perché come detto precedentemente ha ricevuto più lavoro durante l’ultimo periodo e il responsabile è stato particolarmente difficile da rivestire per via dell’inesperienza dei membri a gestire e organizzare un gruppo di persone.

Però nonostante queste difficoltà iniziali a livello organizzativo, causate anche dall'emergenza sanitaria, siamo riusciti a svolgere tutte le attività previste. Noi come gruppo ci sentiamo pienamente soddisfatti del prodotto software e riteniamo di aver svolto un buon lavoro e di aver dimostrato buone capacità di teamworking, sperando di aver soddisfatto le aspettative del proponente.