

Progettazione e sviluppo di uno strumento di classificazione basato su grafi di regole

Andrea Luigi Cambiaghi – 943005

Tesi di Laurea in Informatica - Aprile 2024

Nel campo dell'analisi dati, è fondamentale disporre di uno strumento efficace e intuitivo per gestire la complessità e la varietà delle informazioni a disposizione. Il mio progetto si focalizza sulla realizzazione di un software grafico e interattivo appositamente ideato per la classificazione dei dati, con l'obiettivo di semplificare e ottimizzare il processo analitico. L'idea è permettere agli utenti di creare in modo agevole regole sofisticate per interpretare i propri dataset, assicurando decisioni ben informate e una guida efficace nello sviluppo in diversi settori.

Il focus del progetto è sviluppare un'applicazione web con un editor intuitivo, permettendo agli utenti di creare e manipolare un insieme di regole decisionali attraverso un grafo aciclico direttamente sul browser. Ogni nodo all'interno del grafo rappresenta una condizione booleana, la cui valutazione determina l'arco uscente da seguire nell'analisi dei dati. Le connessioni tra i nodi consentono di collegare le varie condizioni, creando così un flusso logico attraverso il grafo.

Per avere una migliore comprensione del funzionamento dell'applicazione, prendiamo in considerazione un esempio pratico. Supponiamo di avere a disposizione un file un insieme di dati multidimensionali relativi ai clienti di un negozio online, come età, genere, regione di residenza e acquisti effettuati e di avere l'obiettivo di categorizzare i clienti in base a specifiche caratteristiche, come età e genere, per creare profili mirati per strategie di marketing personalizzato.

Attraverso l'applicazione sviluppata, l'utente ha la possibilità di creare un grafo di regole che stabiliscono le condizioni per la classificazione.

Durante il processo di sviluppo di questo strumento, mi sono confrontato con varie sfide tecnologiche e complessità, come la progettazione di un'interfaccia utente intuitiva, l'ottimizzazione delle prestazioni per gestire grandi quantità di dati e la creazione di un ambiente flessibile per consentire agli utenti di personalizzare le regole di classificazione per i propri dati multidimensionali. Nonostante la complessità del processo di sviluppo, il risultato finale è un'applicazione che offre agli utenti un'esperienza di classificazione dati intuitiva e altamente personalizzabile.

Tecnologie coinvolte. Per la mia tesi, ho impiegato BaklavaJS per implementare l'aspetto visuale dell'applicazione. Grazie a questa libreria, ho potuto creare un grafico interattivo che gli utenti sono in grado di personalizzare in base alle proprie preferenze. Per la parte di design, ho integrato HTML e CSS e arricchito alcuni dettagli estetici utilizzando delle librerie aggiuntive.

Ho scelto Flask per la parte di back-end, in quanto questo framework semplifica la comunicazione tra il front end e lo script Python che si occupa dell'applicazione delle regole di classificazione.

Risultati ottenuti. Il lavoro ha portato alla creazione di un sistema avanzato che si basa su grafi di regole per analizzare i dati in modo rapido e comprensibile. Questo nuovo sistema ha dimostrato di accelerare le analisi e migliorare la comprensione dei dati complessi, con possibili utilizzi in diversi settori. Particolare attenzione è stata posta all'usabilità e all'accessibilità, con un'interfaccia utente progettata per consentire a utenti di diverse competenze di sfruttare appieno le potenzialità del sistema.