# Big Data Analytics

**19CSE357**

# Capstone Project

# Blog Content



Team members

## Nimish Kushwaha CH.EN.U4CSE22073

**Aashutosh Kumar pandit CH.EN.U4CSE22076**

## Ayush Kushwaha CH.EN.U4CSE22077

**Abstract**

In the evolving landscape of digital content, the ability to provide personalized recommendations is crucial for enhancing user engagement and satisfaction. This project implements a content-based recommendation system for a blog platform, leveraging machine learning techniques to suggest relevant posts to users based on their reading history. The system employs a combination of **TF- IDF (Term Frequency-Inverse Document Frequency)** and **cosine similarity** to analyze the textual content of blog posts stored in a MongoDB database.

Using the **Flask** framework for the backend, the application connects to MongoDB to retrieve blog post data, which is then transformed into a numerical format suitable for similarity analysis. The **TF-IDF Vectorizer** constructs a matrix that quantifies the importance of terms within the post content, enabling the computation of pairwise cosine similarity scores between posts. This allows the system to identify and recommend the top five posts that are most similar to a given post, enhancing the user experience by presenting content that aligns with individual interests.

The project is hosted on Google Cloud, where users are authenticated for access. When logged in, administrators can delete posts, a feature managed on the server side. When logged out, the system switches to the client side, where regular users can only upload posts and do not have permission to delete them. This ensures that administrative and regular user roles are clearly separated. Additionally, the project uses content-based filtering to provide personalized recommendations, helping to increase user engagement. The interface is user-friendly, and future improvements may include collaborative filtering and advanced natural language processing