SCHOOL OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE (SAIDE)

BIG DATA MANAGEMENT

PROJECT PROPOSAL

Under Supervision

Of

PROF. DIP SANKAR BANERJEE

SUBMITTED BY:

Naresh G (G23AI2011)

Saif Jawed (G23AI2026)

Sudherson G (G23AI2125)

***Title*:** ShopSmart - Personalized Recommendation using Big Data Analytics.

Problem Statement: The "ShopSmart" project's objective is to create a strong recommendation system that uses big data analytics to deliver real-time, individualized product recommendations. In order to improve the shopping experience, boost engagement, and boost revenues, the system analyzes demographic information, past purchases, and customer behavior. By providing personalized recommendations that appeal to each unique consumer, this project aims to alleviate the problem of information overload, eventually promoting loyalty and enhancing overall corporate performance.

In the competitive retail landscape, businesses struggle to meet the diverse preferences and expectations of their customers. Traditional marketing approaches often fail to deliver personalized shopping experiences, leading to lower customer satisfaction and retention rates. Consumers are inundated with options, making it challenging for them to find products that align with their interests and needs.

List of Key Technology Challenges:

Data Processing at Scale: Handling large amounts of customer interaction data, including clickstreams, purchase histories, and real-time activity.

Accurate Personalization: Developing models that can provide highly personalized and relevant recommendations based on individual behaviour patterns.

Real-time Analytics: Offering recommendations in real-time, without latency, for a smooth and seamless customer experience.

Diverse Data Sources: Incorporating structured (transaction data) and unstructured data (reviews, social media, etc.) into the recommendation engine.

Dynamic Adaptation: Ensuring that the recommendation engine adapts quickly to changing user preferences and behaviour.

Technology Stacks involved:

- Cloud Infra – GCP/ AWS
- Data collection – Datalakes
- Workflow/ Data Pipeline – Airflow
- Data Storage - S3, GCS (Google Cloud Storage), NoSQL
- Data processing/Warehousing – Apache Spark/ BigQuery
- Dev. Environment - Flask.
- Scalability – Docker, GKE (Google Kubernetes Engine

GCP act as a PaaS environment where we can store, compute and process of large volume, Velocity and variety of the data.

Codebase – Github as a repository, as well as development lifecycle.

Deployment - Google Cloud Platform.

Benchmarking - Amazon analytics *vs* Shop smart.

API – Flask - Python Framework