Saarthak Gupta

linkedin | Github | Medium

Email : Saarthak.gupta.mec22@itbhu.ac.in
Mobile : 7303277192

EDUCATION

- **Indian Institute of Technology(BHU)** — Varanasi, India
  Integrated Dual Degree in Mechaincal Engineering; CGPA: 8.33 — 2022-2027
- **S.R.D.A.V. Public School** — Delhi, India
  CBSE(Class XII); CGPA: 93.8% — 2022

SKILLS

- **Languages:** Python(pandas, numpy, matplotlib, seaborn) ,sql
- **Frameworks:** scikit learn, tensorflow, keras , Langchain
- **Tools:** git, Mysql, excel, powerbi
- **Platforms:** windows, aws

PROJECTS

CLV PREDICTION FOR AUTO-INSURANCE COMPANY | LINK

Developed a comprehensive Customer Lifetime Value (CLV) prediction model to improve customer retention and marketing strategies, leading to more informed business decisions.

- Leveraged machine learning techniques to predict CLV, optimizing hyperparameters with GridSearchCV for accurate predictions.
- Conducted comprehensive data preprocessing followed by detailed EDA using statistical summaries, distribution analysis, and visualizations to uncover insights.
- Created a web interface and dashboard for CLV predictions and customer insights, and implemented a Q&A platform using Google's Gemini LLM to facilitate natural language queries, enhancing marketing strategies and customer retention efforts.

REAL ESTATE PRICE PREDICTION MODEL | LINK

Objective was to build a complete product which can be used by any end user who wishes to buy property

- Used Random Forest Regressor trained on 40,000+ data points, achieving R^2 score of 0.90 and MAE of 0.45.
- Implemented using TF-IDF vectorizer and cosine similarity, providing weighted recommendations based on various features.
- Analytics module provides the valuable insights on existing properties of Gurgaon using analytical libraries like plotly and seaborn.

MEDICAL CHATBOT | LINK

Objective was To develop a medical chatbot capable of answering user queries related to medical information and providing relevant resources in a timely and accurate manner, utilizing natural language processing and information retrieval techniques.

- Extracted and preprocessed text from medical PDFs using Langchain.
- Used Hugging Face embeddings and Pinecone for efficient similarity search.
- Developed a custom PromptTemplate and used CTransformers model for conversational responses. Created a user-friendly interface for interaction.

HONORS AND AWARDS

- Participant in Convolve (PAN IIT Data Analytics event)
- Second runner up in statistella