GAURI TRIPATHI

gauritripathi2002@gmail.com LinkedIn Github

EDUCATION

SRM University, KTK **Sept '20 – June'24**

B.Tech in Computer Science Engineering *9.02 CGPA*

EXPERIENCE

Surya Consultants **Jun '23 – Dec'23**

Data Scientist

- Developed and implemented a dynamic client data analysis system using Python and SQL, leading to a 30 percent improvement in data processing efficiency.
- Employed ML models like Logistic Regression for predictive modeling and trend analysis, resulting in a better analysis accurate market forecasts for clients.
- Collaborated in the creation of an interactive dashboard using Plotly, Seaborn for real-time data visualization, enhancing client presentations and decision-making processes.
- Conducted comprehensive data cleansing using SpacyNLP, ensuring a increase in data readability.

PROJECTS

Doctor LLM | *Transformers, bitsandbytes, Torch* | Github

- Advanced AI Model: Built on Mistral-7B, a powerful model that can generate thoughtful and accurate medical advice.
- Quantized for Efficiency: Future enhancements include quantizing the model to 2, 4, and 8 bits for task-specific use cases, optimizing performance without compromising accuracy.
- High-Quality Interaction: Ensures factual and nuanced communication, prioritizing user well-being and understanding.
- Logical / IQ Tasks: Perform roleplays, and chain-of-thought reasoning.

COVID-19 Data Analysis Project | *SQL, Apache Superset* | Github

- comprehensive analysis of COVID-19 data, focusing on various aspects like death rates, infection rates, and vaccination progress across different countries and continents
- Apache Superset Integration: these queries will be implemented on SQL Apache Superset to enhance data visualization and dashboarding capabilities

DeepTrafficPredictor | *LSTM, NumPy, skelarn, pandas* | Github

- Data Preprocessing and Preparation: using Numpy and Pandas, involved in loading and normalization of traffic data through Min-Max scaling, ensuring data readiness for model input.
- LSTM Model Training: an LSTM-based Sequential NN, integrating Dense layers and leveraging the Adam optimizer. trained the model over 10 epochs, focusing on loss reduction for enhanced model accuracy.
- Model Evaluation and Data Visualization: calculating RMSE for both training and testing datasets. utilized Matplotlib for data visualization, effectively comparing actual vs. predicted traffic counts to demonstrate model performance..

Starbucks Nutritional Analysis | *Pandas, Matplotlib, Seaborn* | GitHub

- Data Loading, Preprocessing, and Exploration: using Pandas, loading and exploring the Starbucks dataset. Conducted thorough initial data analysis and optimized dataset
- Statistical Analysis and Data Cleaning: extracting key insights like mean, standard deviation, and quartiles
- used Matplotlib and Seaborn for creating insightful visualizations including scatter plots, bar plots, and histograms to elucidate relationships between various nutritional elements.

TECHNICAL SKILLS

Languages:Python, SQL
Technologies: Anaconda, Apache, Linux, Jupyter, PostgreSQL, Azure, Shopify, WordPress
https://www.overleaf.com/project/65bbbd270b3f5483d56eb255 **Developer Tools**: Git, Docker, Microsoft Azure ML, VS Code, AWS Sagemaker, AWS EC2
Libraries: PyTorch, Torch, Langchain, chromaDB, Numpy, Pandas, Scikit-Learn, OpenCV, FastAPI, FastAPI
MLOPS: WeightsBiases

EXTRA-CURRICULAR

- AMVMUN 2018, 2019: Special Mention in the AMVMUN for the All India Political Party Meet.
- Sub Editor Photographer: Attended the AMVMUN as part of the Photography Committee