Yash Soni

📞 +91 7389296458 ✉ yashmukeshsoni@gmail.com in LinkedIn ⬡ Github

Technical Skills

Languages: Python, Postgre-SQL

Packages and Modules: Numpy, Pandas, MatplotLib, Seaborn, Scikit-Learn

Tools and Technology: AWS, Azure-Cosmos, Power-BI, MS-Excel, Tableau, GitHub, VS-Code, Google-Colab, Jupyter Notebook

Algorithms: Hugging face Transformers, Linear/Logistic Regression, Random Forest, DecisionTree, SVM, KNN, Neural-Networks, XGBoost, K-Means Clustering, NLP, Recommender-System

Professional Skills: Chat-GPT Prompt Engineering, Data Analytics, Data Visualization, Data Wrangling, DataCleaning, Machine Learning Modeling

Personal Projects

End-to-End NLP Project: Text Summarizer.   Source Code

- Utilized **Hugging Face Transformers** library, leveraging pre-trained model **google/pegasus-cnn-dailymail**, to expedite development and enhance summarization accuracy, within an organized project structure for efficient code management and clarity.
- Devised a **Modular Pipeline** comprising data ingestion, validation, transformation, model training, and evaluation stages, implementing **Logging Mechanisms** throughout for comprehensive tracking of execution steps, aiding in debugging and monitoring progress.
- Employed modular coding and **Object-oriented programming (OOP)** principles to enhance code readability, reusability, and maintainability, while leveraging **YAML configuration** files for**Continuous Integration/Continuous Deployment (CI/CD)** workflows, ensuring streamlined automation of build, test, and **Deployment** processes.
- Implemented a **RESTful API** using **FastAPI**, enabling user interaction with the summarization system while employing ROUGE metrics (**ROUGE-1, ROUGE-2, ROUGE-L, ROUGE-L-SUM**) for assessing summarization model performance, indicative of achieving satisfactory summarization quality.
- Leveraged **GitHub Actions** for version control automation, automating tasks like testing, linting, and deployment based on predefined triggers, while **Dockerizing** the application for seamless deployment, packaging it with dependencies into containers for efficient **Deployment** on **AWS ECR**, ensuring scalability and reliability in **Cloud environments**.

Internship Experience

Factacy.AI   Mar 2024 – Jun 2024
Data Science Intern   *Gurugram, HR*

- Utilized in **Python** for data manipulation, analysis, and visualization, contributing to innovative AI-driven solutions.
- Employed **Excel** for data preprocessing and analysis, enhancing data quality and facilitating efficient decision-making processes.
- Managed and optimised data storage using **Azure Cosmos** and leveraging **NoSQL** databases for scalable storage and retrieval. Applied **Machine Learning** techniques to derive insights, collaborated with **NLP** and Computer Vision experts, and contributed to revitalizing Private Equity and Venture Capital markets through augmented data analytics. Developed bespoke **AI solutions** for startups worldwide to monetize data and drive growth

Relevant Courses

Full Stack Data Science and AI   Nov 2022 – July 2023
Cohort Durban   *AlmaBetter (Bengaluru, Karnataka)*

- Completed Full Stack Data Science and AI Course at AlmaBetter, mastering **Python, Machine learning, Power BI, Excel, and Postgre-SQL**.
- Proficient in Python programming, machine learning algorithms, Power BI data visualization, advanced Excel techniques, and Postgre-SQL database management for comprehensive data analysis and AI solutions

Education

Ujjain Engineering College, Ujjain   **2019 - 2022**
B.Tech in Civil Engineering   *7.48/10 CGPA*

Algorithmic Competitions/Achievements

- Earned **Python 5-Star Gold Badge** on HackerRank Link.
- Earned **SQL 5-Star Gold Badge** on HackerRank Link.