Pratyush Srivastava
Data Scientist and Analytics Engineer

+91 9721373252 | pratyushsrivastava500@gmail.com | Linkedin | Portfolio | Github

PROFESSIONAL SUMMARY

Results-driven Data Scientist with 3+ years of experience in Machine Learning, Deep Learning, NLP, and Generative AI. Proficient in developing end-to-end data-driven solutions, with expertise in LLMs (OpenAI, Vertex AI), LangChain, NLTK, Scikit-learn, TensorFlow, Keras, NumPy, and Pandas. Skilled in data analysis, visualization, and feature engineering to extract actionable insights. Experienced in cloud computing (GCP) and deploying high-performance FastAPI applications. Adept at collaborating cross-functionally to drive business impact through AI-driven innovation.

EDUCATION

Galgotias University Greater Noida, Uttar Pradesh
Bachelor of Technology in Computer Science and Engineering(Major GPA:8.78) *Aug. 2018 - May 2022*

Green Field Academy Lakhimpur Kheri, Uttar Pradesh
Intermediate, PCM(84.8) *April. 2017 - May 2018*

Green Field Academy Lakhimpur Kheri, Uttar Pradesh
High School(Major CGPA:10.0) *April. 2015 – May 2016*

EXPERIENCE

Digital Specialist Engineer(Analytics) July 2022 – Present
Infosys Limited *Chandigarh,Punjab*

- Collaborated with cross-functional teams to define project requirements, ensuring AI solutions and analytics meet organizational goals.
- Partnered with project managers to identify analytics needs, establish critical metrics and KPIs, and provide actionable insights to drive informed decision-making.
- Designed, developed, and optimized generative AI models to improve performance, scalability, and efficiency in alignment with business objectives.
- Developed and maintained AI and data pipelines, handling data preprocessing, feature extraction, model training, and evaluation to ensure efficient data flow.
- Contributed to the development of best practices for generative AI and data systems, troubleshooting and debugging issues while embracing new technologies to enhance analytical outcomes.

PROFESSIONAL PROJECTS(RECENT)

LGE US Sales and Site Analytics Assistant | *Python , NLP , SQL , VertexAI , BigQuery*

- ∗ Collaborated with clients to gather requirements and design a highly efficient sales analytics chatbot solution, leveraging natural language processing (NLP) for enhanced user interaction.
- ∗ Engineered a seamless integration of the sales analytics chatbot with Vertex AI and Gemini LLM model, achieving real-time data processing that enhanced user query response times by an average of 3 seconds.
- ∗ Designed and developed a robust query framework to construct and execute SQL queries, streamlining data retrieval and improving user insights through automated summarization.
- ∗ Implemented an NLP pipeline to extract and map user entities, increasing accuracy in query interpretation and conversational responses.
- ∗ Converted user queries into SQL statements, transforming data into natural language insights and delivering dynamic visualizations to enhance decision-making in a conversational format.

Home Seller Assistant | *Python , Large Language Models(LLMs), RAG , LangChain , GCP*

- ∗ Leveraged large language models (LLMs) to develop a scalable solution for automating data extraction from mortgage documents within an GCP cloud environment.
- ∗ Implemented a Retrieval-Augmented Generation (RAG) approach to classify and extract relevant information from mortgage document streams, streamlining the data population process into the user interface (UI).
- ∗ Developed a Conversational AI solution, incorporating a document QnA chatbot powered by enterprise-grade vector databases and document embeddings using LangChain, delivering an efficient and intelligent document processing workflow.

Personal Projects

Medical Insurance Price Prediction | *Python , Predictive Modeling , Pandas , NumPy*
- Designed a regression-based model to accurately predict insurance premiums using personal and health data, including age, gender, BMI, smoking status, and region.
- Utilized Pandas and NumPy for efficient data manipulation and analysis, ensuring robust dataset preparation.
- Leveraged Matplotlib and Seaborn to create detailed visualizations, identifying key factors influencing insurance premiums.
- Implemented the model to provide accurate premium predictions, enabling data-driven insurance pricing strategies and supporting business decisions.

Mushroom Classification Web App | *Python, Machine Learning, Streamlit, Data Classification*
- Developed a classification-based model to predict whether a mushroom is edible or poisonous based on its attributes such as color, size, and shape.
- Applied various machine learning algorithms to analyze the dataset and achieve high classification accuracy.
- Utilized Pandas and NumPy for efficient data preprocessing and feature engineering.
- Integrated the model with Streamlit to create an interactive web application for real-time predictions.
- Demonstrated expertise in machine learning, data classification, and web deployment through an intuitive and user-friendly interface.

Technical Skills

Languages: Python, C, SQL (MySQL, PostgreSQL), HTML
Cloud Technologies: Google Cloud Platform (GCP), AWS, Heroku
Developer Tools: Git, VS Code, PyCharm, IntelliJ, Eclipse, Anaconda
Libraries and Frameworks: Pandas, NumPy, Matplotlib, Scikit-learn, TensorFlow, Keras, NLTK, Langchain
Software Tools: MS Word, PowerPoint, MS Excel, Power BI
Technologies: Data Analysis, Machine Learning, Deep Learning, Data Engineering, Web Scraping
Other Skills:Business Analytics, Data-Driven Decision Making, Statistical Analysis, Customer Insights, Predictive Modeling,Model Deployment

RESEARCH AND PUBLICATIONS

Attendance Management System Using OpenCV: Published in the 3rd IEEE International Conference on Advances in Computing, Communication, Control, and Networking (Scopus Indexed), March 2022.
Food Calorie Prediction Using MobileNet: Selected for presentation at the International Conference on Advanced Research and Innovation (ICARI-2022).

CERTIFICATIONS AND ACHIEVEMENTS

All India Rank 6 in National Engineering Olympiad 2021(NEO 4.0)
Finalist of Ami-Techfiesta 2020(A Programming Challenge).
Cleared HackwithInfy All India Coding Contest of Infosys in 2021.
Microsoft Azure AI Fundamentals Certified.
Microsoft Technology Associate Certified in Python Programming.
Infosys Certified Python Programmer.
Infosys Certified Python Associate