Pankaj
Data Scientist

Agra, U.P., India   +91 8630795434   ✉ pankajshakya627@gmail.com   [in] linkedin.com/in/pankajshakya627

○ github.com/pankajshakya627

PROFILE

Data Scientist with nearly 4 years of experience in natural language processing and machine learning model optimization. Skilled in deploying solutions using **Python, Flask, and AWS,** leading to improvements in predictive analytics and user satisfaction. Focused on leveraging data insights to drive business growth and enhance customer experiences.

EXPERIENCE

ITC Infotech - Gurugram, Haryana, India   February 2024 – Present

Data Scientist

- Spearheaded an interdisciplinary team of 3 to architect advanced **machine learning models**, boosting predictive analytics accuracy by 20%.
- Streamlined data science processes leveraging **Scikit-learn**, slashing analysis time by 30%.
- Shaped product development strategy by delivering pivotal data-driven insights on customer behaviour.
- Deployed Dash and **SQL** to craft interactive data visualizations, demystifying complex data for stakeholders and enhancing decision-making processes.

Jasper Colin - Noida, U.P., India   September 2023 – February 2024

Machine Learning Engineer

- Executed extensive analysis on financial news and company data for due diligence, delivering comprehensive insights into company performance and stock portfolio trends.
- Orchestrated the development and **fine-tuning of an LLM** (LLAMA 2) using PEFT, focusing on text summarization of financial news to detect trends and patterns in specific companies.
- Devised scalable solutions utilizing **LangChain** and **FastAPI**, enhancing the extraction and summarization of financial insights.
- Launched the model on an **AWS EC2** instance, ensuring reliable and efficient access to the summarized financial data.

LTIMindtree - Bengaluru, Karnataka, India   March 2021 - August 2023

Python Developer

- Collaborated on the development of over 15 innovative features using Python, Flask, and REST API.
- Developed Python documentation sites, reducing search time by 10%, demonstrating expertise in Git, GitHub, and various cloud tools.
- Facilitated the creation of a data visualization dashboard using Dash, significantly enhancing data interpretation for stakeholders.

EDUCATION

Rajkiya Engineering College — Ambedkar Nagar, U.P.   July 2016 - September 2020
Bachelor of Technology in Information Technology
Major in Computer Science, Statistics & Machine Learning

Maharaja Agrasen Inter College — Agra, U.P.   July 2014 - June 2015
Intermediate - Mathematics and Science

SKILLS

- **Programming Languages**: **Python**, R, SQL, C++
- Data Analysis: Pandas, NumPy, Scikit-learn, SciPy, Statsmodel
- **Machine Learning**: TensorFlow, Keras, PyTorch, NLTK, Gensim, RNN, LSTM, CNN, OpenCV
- **Generative AI (GenAI)**: **Large Language Models (LLMs)**, **Langchain, Streamlit, transformers, GenAI with structured and unstructured data,**
- **Data Visualization**: Matplotlib, Seaborn, Plotly, Dash
- **Databases**: MySQL, PostgreSQL, MongoDB
- **Other Tools: Git, Docker**, Databricks, Kubernetes **AWS, GCP, AWS SageMaker**, AWS Jumpstart, Jupyter Notebook, Excel, Scrapy, BeautifulSoup, Linux

PROJECTS

Fine-Tuning a Large Language Model
- Fine-tuned the LLAMA3-8B model using Parameter-Efficient Fine-Tuning (PEFT) techniques, significantly reducing computational overhead and improving task-specific performance
- Achieved substantial performance improvements without extensive computational resources, enhancing the sustainability and cost-effectiveness of the fine-tuning process.

AI-PDF Assistant
- Integrated OpenAI and HuggingFace APIs: Developed a Python-based PDF comprehension tool, combining OpenAI's and HuggingFace's APIs with Langchain for accurate, real-time query responses.
- Enhanced User Interaction: Crafted a Streamlit interface for improved user experience and employed PyPDF2 for efficient PDF data handling.

Sentiment Analysis Web Application
- Engineered and deployed a sentiment analysis web application on AWS EC2 using Python and Flask, achieving an 85% accuracy rate in analyzing over 10,000 tweets.
- Utilized scalable algorithms and integrated Twitter API for real-time data collection, processing 5,000+ tweets to enhance customer sentiment understanding for targeted marketing.

Find Other Project Here 🔗

OTHERS

Certifications:

- Generative AI with **Large Language Models**   - DeepLearning.AI and AWS   🔗
- Python for **Data Science**, AI & Development   - IBM   🔗
- Applied Plotting, Charting & Data Representation in Python Data Science Methodology - IBM   🔗
- The Top 5 Machine Learning Libraries in Python   - Udemy   🔗
- Complete Python programming from Basics to Advance level   - Udemy   🔗
- National Workshop on "Image Processing and Computer Vision & Application"

Awards:
- Strategic Contributor – LTIMindtree (2022)
- Rising Star Award – LTIMindtree (2021)