Pankaj
Data Scientist

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

PROFILE

Innovative Data Scientist with nearly 4 years of experience specializing in natural language processing and machine learning model optimization. Proven expertise in deploying scalable solutions using Python, Flask, and AWS, driving significant improvements in predictive analytics and user satisfaction. Passionate about leveraging data insights to support business growth and enhance customer experiences

EXPERIENCE

ITC Infotech - Gurugram, Haryana, India   *February 2024 – Present*
Data Scientist
- Led an interdisciplinary team of 3 to develop advanced machine learning models, enhancing predictive analytics accuracy by 20%.
- Optimized data science processes using Scikit-learn, reducing analysis time by 40%.
- Implemented a deep learning model that improved the natural language processing capabilities of the company's software, leading to a 15% increase in customer satisfaction
- Influenced product development strategy by providing key insights from data analytics, resulting in a 20% increase in revenue
- Successfully utilized Tableau and SQL for data visualization, making complex data understandable for stakeholders and enhancing decision making process

Jasper Colin - Noida, U.P., India   *September 2022 – February 2024*

Machine Learning Engineer
- Conducted extensive analysis on financial news and company data for due diligence, delivering comprehensive insights into company performance and stock portfolio trends.
- Led 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.
- Implemented scalable solutions using LangChain and FastAPI, enhancing the extraction and summarization of financial insights.
- Deployed the model on an AWS EC2 instance, ensuring reliable and efficient access to the summarized financial data.

LTIMindtree - Bengaluru, Karnataka, India   *March 2021 - August 2022*
Python Developer
- Collaborated on the development of over 15 innovative features using Python, Flask, and REST API.
- Supported the development and Reduced search time by 10% through the creation of Python documentation sites, demonstrating expertise in Git, GitHub, and various cloud tools.
- Assisted in the creation of a data visualization dashboard using Dash that improved data interpretation for stakeholders

EDUCATION

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

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

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:** Large Language Models (LLMs), Langchain, Streamlit
- **Data Visualization:** Matplotlib, Seaborn, Plotly, Dash
- **Databases:** MySQL, PostgreSQL, MongoDB
- **Other Tools:** Git, Docker, AWS, GCP, AWS SageMaker, AWS Jumpstart, Jupyter Notebook, Excel, Scrapy, BeautifulSoup

PROJECT

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
- **High-Availability Deployment:** 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.
- **Efficient Data Processing:** Utilized scalable algorithms and integrated Twitter API for real-time data collection, processing 5,000+ tweets to enhance customer sentiment understanding for targeted marketing.

OTHER

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:
- Impact Team Player – LTIMindtree (2022)
- Rising Star Award – LTIMindtree (2021)