Shubham Pandey

Mobile: +91-8756169800 E-mail: shubham30p@gmail.com Github : /shubhamp1028 LinkedIn: /in/shubham1028

Profile

Data Science undergraduate (Batch 2026) with hands-on experience in Python, SQL, ML/AI model development, and data visualization. Skilled in ETL pipelines, feature engineering, and deploying ML applications using Flask, Streamlit, and Azure. Currently gaining hands-on experience in PySpark for scalable data processing, with an applied project on E-commerce Customer Sales Data Analysis. Passionate about solving real-world problems using data, with a strong foundation in statistics, machine learning, and business-oriented analysis.

Skills

- **Languages:** Python, PyTorch, TensorFlow, Scikit-learn, Hugging Face, Transformers, CNNs
- **ETL & Database:** MySQL, SQL, Data Cleaning, Preprocessing, PySpark
- **Tools & BI:** Tableau, Power BI, Matplotlib, Seaborn, Pyplot, EDA, Feature Engineering, Time Series
- **DevOps :** Flask, Streamlit, Docker, Azure, Git/Github
- **Soft Skills:** Analytical Thinking, Stakeholder Collaboration, Communication, Problem-Solving

Projects

Customer Sales Analysis | *PySpark, SQL, Databricks*

- Building an ETL pipeline to load, clean, and analyze simulated e-commerce sales data.
- Using PySpark to perform aggregations, joins, and transformations on customer orders.
- Goal: Provide insights into sales trends, repeat customers, and revenue growth opportunities.

Plant Disease Detection | *Tensorflow, CNN, Azure* | Github

- Trained a CNN classifier achieving 92% accuracy across multiple plant disease categories.
- Deployed model on Azure Docker Container, enabling real-time image predictions in ~5 seconds.
- Integrated Grad-CAM visualizations to improve model explainability for end users.

Resume–Job Match Detector | *BERT, PyTorch, NLP* | Github

- Fine-tuned BERT NER model on 25K+ resume tokens, improving entity recognition by 15% with weighted loss.
- Built a resume–job similarity pipeline that increased matching accuracy from 70% to 83%.
- Packaged as a prototype tool to streamline candidate–JD shortlisting.

DocuPi: Document Chat | *Vector DB, Langchain, Gemini API* | Github

- Developed a retrieval-augmented generation chatbot for policy documents with LangChain.
- Enabled 90%+ accuracy in document Q&A retrieval by integrating embeddings + vector databases.
- Designed modular architecture for scalable deployment across multiple document types.

Focus & Posture Guide | *OpenCV* | Github

- Developed a retrieval-augmented generation chatbot for policy documents with LangChain.
- Enabled 90%+ accuracy in document Q&A retrieval by integrating embeddings + vector databases.
- Designed modular architecture for scalable deployment across multiple document types.

Internships

BearHugs - Data Analysis Intern | Hyderabad June, 2025-Aug, 2025

- Built interactive Tableau dashboards to present key KPIs to stakeholders.
- Automated data pipelines to reduce reporting time by ~30%.
- Analyzed business datasets and provided actionable insights for risk assessment.

Education

Lovely Professional University August, 2022 - July, 2026

Bachelor in Technology , Data Science and Machine Learning

CGPA: 7.51

Sunbeam School April, 2017 - March, 2021

CBSE – XII (PCM) – 91.8%

CBSE – X – 91%

Certification

Business Intelligence Analyst Professional | IBM December, 2024

Power BI Data Analyst Professional | Microsoft November, 2024

AI/ML Engineer Professional | Microsoft August, 2025

Data Warehouse Engineer | IBM June, 2025

Data Engineering | GeeksforGeeks July, 2025