Shubham Pandey

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

Profile

Aspiring Data Scientist and Machine Learning enthusiast with hands-on experience in predictive modeling, deep learning, and data visualization. Proficient in Python, scikit-learn, TensorFlow, and PyTorch, with practical exposure to projects in stock price prediction using LSTM, crop yield and climate analysis, and exoplanet discovery. Skilled in translating complex datasets into actionable insights and building end-to-end ML pipelines, including model deployment. Passionate about leveraging AI and analytics to solve real-world problems and contribute to data-driven decision-making.

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.

Stock Price Prediction | *LSTM, Statistical Analysis, Tableau* | Github
- Built an LSTM-based model to forecast stock price movements from time-series data, R^2 score of ~0.82 on test data.
- Implemented data preprocessing pipelines (normalization, sequence generation) in Python with Scikit-learn, improving model accuracy by 15% over baseline approaches.
- Designed interactive Tableau dashboards to visualize stock trends, predicted vs. actual performance, and key financial indicators, enabling clear insights for decision-making.

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