HARSHIT JAIN

hjain0516@gmail.com | +91 8085534588 | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Aspiring Data Analyst/Engineer skilled in Python, SQL, and end-to-end data workflows. Experienced in building dashboards, optimizing queries, automating ETL processes, and developing ML models that support product decisions. Strong foundation in statistical analysis, time-series modelling, and transforming data into actionable insights.

SKILLS

- **Analytics:** Exploratory Analysis, Statistical Methods , Data Visualization, Time-Series Analysis
- **Machine Learning:** Regression, Classification, Model Evaluation, Feature Engineering, CNNs, Transfer Learning
- **Data Engineering:** ETL Pipelines, SQL Optimization, Data Modeling, API Development, Workflow Automation
- **Languages:** Python, SQL (Joins, CTEs, Window Functions), JavaScript, C++, MySQL
- **Frameworks/Libs:** Pandas, NumPy, Scikit-learn, TensorFlow, Keras, Streamlit, Plotly
- **Tools:** Power BI, Tableau, Excel, Git/GitHub, VS Code, Jupyter, Google Colab

EXPERIENCE

Bluestock Fintech **Nov 2024 – Dec 2024**
Software Development Engineer Intern *Remote, Pune*

- Improved dashboard query performance by **20%** by restructuring PostgreSQL workflows and removing redundant joins, enabling faster analytics for a **500K+ user base**.
- Designed and deployed **REST APIs** that standardized internal data access and reduced data retrieval latency by **30%**.
- Automated ETL pipelines for log ingestion, reducing manual processing time and enabling timely weekly business insights.

PROJECTS

Stock Market Analytics Platform & Sector Risk Modeling
Python, Pandas, NumPy, Streamlit, Plotly, yfinance

- Engineered a modular ETL pipeline processing **500K+ stock records** with automated ingestion, cleaning, sector-wise aggregation, and feature engineering.
- Built comprehensive analytics dashboard using Python (Streamlit, Plotly) featuring **15+ interactive visualizations** for sector performance, risk metrics, and cross-correlations, processing data for 50+ stocks with less than 2-second load time.
- Implemented a time-series modelling layer (Sharpe Ratio, Max Drawdown, Rolling Volatility), identifying the **IT sector as strongest risk-adjusted performer**.
- Optimized dashboard performance to **less than 2 seconds load time** using caching, column pruning, and efficient in-memory structures.
- Discovered insights such as Banking–Realty **correlation (0.81)** and IT–Energy **inverse correlation (-0.23)**, supporting diversification strategies.

Pneumonia Detection from Chest X-rays using Deep Learning
Python, TensorFlow, Keras, OpenCV, Transfer Learning

- Trained a CNN model on **5,856 chest X-rays**, using InceptionV3 transfer learning and a custom deep CNN architecture, achieving **89.53% accuracy**, **95.48% recall**, outperforming baseline models.
- Improved accuracy with augmentation, regularization, and transfer learning (InceptionV3), reaching **92% validation accuracy**.
- Built a Streamlit app for real-time uploads, predictions, confidence scoring, and Grad-CAM interpretability.

EDUCATION

Vellore Institute of Technology (VIT) ***Sept 2022 – May 2026***
B.Tech in Computer Science with AI & ML *CGPA: 7.9*

CERTIFICATIONS & ACHIEVEMENTS

- Coursera: Applied Machine Learning in Python
- NPTEL: Privacy & Security in Online Social Media
- FacePrep: MongoDB Associate Database Administrator
- VIT: Python Essentials, Computer Vision
- **Karnataka Police Datathon (Semi-Finalist):** Built a YOLO-based congestion & encroachment detection system.
- **FinTech Club (PR & Outreach):** Led partnerships and community growth initiatives.