

# Fahad Shaikh

New York City — fs12516n@pace.edu — +1 (329) 204 9984  
LinkedIn — GitHub

## Summary

Highly accomplished Data Scientist with 4+ years of experience delivering impactful ML/DL solutions for Fortune 100 clients, including Microsoft, Citigroup, LSEG, and Walmart. Expertise spans developing NLP applications, time series forecasting, and business intelligence tools, driving actionable insights from complex datasets. Proficient in statistical modeling and deep learning frameworks, with a strong commitment to staying at the forefront of AI advancements and contributing to academic research.

## Skills

- **Programming & Machine Learning:** Python, R, SQL, Scikit-learn, TensorFlow, Keras, PyTorch, XGBoost, Pandas, NumPy, OpenCV, LangChain
- **Deep Learning:** ANN, CNN, RNN, Transformers, Autoencoders, LSTMs, BERT, GPT-4 (API)
- **NLP & Applied AI:** Topic Modeling, Named Entity Recognition, Question Answering, RAG Pipelines, Sentiment Analysis
- **Cloud & Databases:** Azure, AWS (S3, Redshift, EMR), Oracle, Snowflake, Azure Databricks
- **Visualization:** Power BI, Tableau, Matplotlib, Plotly, ChromaDB
- **DevOps:** Docker, Kubernetes, Git, CI/CD Pipelines, JIRA

## Education

- Pace University, New York City** *Sept 2024 – Jan 2026*  
MS in Data Science (GPA: 4.0)  
Relevant Coursework: Deep Learning, Scalable Databases, AI, Data Mining, Business Data Science
- Savitribai Phule Pune University, India** *Aug 2014 – May 2018*  
BE in Electronics and Telecommunication (GPA: 7.0)

## Professional Experience

- Data Scientist** *Nov 2023 – Aug 2024*  
*Endeavor Group (IMG Media), New York*
- **Enhanced** HR retention by 15% by building predictive churn models utilizing neural networks and Cox models.
  - **Developed** NLP pipelines to classify attrition reasons from resumes and exit surveys, providing key insights for HR strategies.
  - **Designed** and **deployed** interactive dashboards using Power BI, leveraging Azure SQL and Azure Databricks for real-time HR analytics.
  - **Orchestrated** ETL pipelines and automated model scoring using Azure Data Factory.
- Sr Data Consultant** *Apr 2019 – Jun 2022*  
*Mu Sigma, Bengaluru*
- **Led** customer segmentation for Microsoft Viva using K-means, GMM, and hierarchical clustering, **identifying** key user groups.
  - **Engineered** anomaly detection systems with Isolation Forest, One-Class SVM, and LOF for Citigroup, **improving** fraud detection capabilities.
  - **Automated** LSEG trading report generation using Python, Tableau, and Oracle, **reducing** manual effort by 150 hours monthly.
  - **Built** and **deployed** seasonal forecasting models using ARIMA and Prophet for improved business planning.
  - **Utilized** AWS EMR, S3, EC2, and Redshift to **enable** scalable ML training and distributed analytics for various client projects.
- Trainee** *Aug 2018 – Mar 2019*  
*KPIT, Pune*
- **Developed** adaptive cruise control logic using embedded C and PID control.
  - **Managed** automotive ECU requirements using IBM DOORS.

## Projects

---

### Academic Paper RAG Tool (Jan 2025)

- **Engineered** a citation-aware insight extraction tool leveraging LangChain, GPT-4, and ChromaDB.
- **Reduced** retrieval latency by 40%, significantly enhancing academic research workflows.

### WebChat Assistant for Websites (Dec 2024)

- **Developed** an intelligent chatbot using OpenAI and BeautifulSoup for enhanced website interaction.
- **Achieved** a 40% improvement in UX quality and a 60% reduction in query latency.

### Flight Delay Prediction Model (Oct 2024)

- **Built** a flight delay classification model achieving an AUC of 0.84 using Random Forest and XGBoost.
- **Implemented** scalable retraining by storing FAA datasets in AWS S3 and triggering EMR jobs.

## Achievements

---

- **Impact Award**, LSEG: **Redesigned** revenue-risk model, resulting in \$400K annual savings.
- **Spot Awards**, Mu Sigma: **Delivered** critical solutions for Microsoft and Citigroup, recognized for exceptional contributions.
- **Participation**: Smart India Hackathon, 2017.

## Certifications

---

- DeepLearning.AI TensorFlow Developer — Coursera
- HuggingFace NLP Specialization
- Google Advanced Data Analytics — Coursera
- AWS Certified Machine Learning – Foundational

## Interests

---

- NLP Research, Academic Writing, Open Source Contributions, Tech Blogging, Mentoring