Akshit Sanoria

+1 (480)-790-0074 |armdoor745@gmail.com |linkedin.com/in/akshit-sanoria/|github.com/Armdoor

PROFESSIONAL SUMMARY

Computer Science graduate (May 2025) with hands-on experience in end-to-end data science lifecycles, ML pipeline development, LLM fine-tuning, and scalable data workflows on AWS. Proficient in Python, PySpark, REST API design, SQL/NoSQL modeling, CI/CD automation, and NLP, with a solid foundation in data analysis and machine learning.

EDUCATION

B.S. in Computer Science, Minor in Human Development **August 2022-May 2025**
University of Maryland, College Park College Park, Maryland
Relevant Coursework: Object-Oriented Programming; Data Structures; Statistics; Economics; Systems Programming in C; Algorithms; Data Science; Machine Learning; Natural Language Processing; Computer Vision

WORK EXPERIENCE

Machine Learning Intern **August 2024-November 2024**
Tekizma

- Built a custom text scraping tool from scratch to automate data extraction, improving data collection efficiency by 60% for downstream ML pipelines.
- Processed large real-world datasets (300MB-5GB) and implemented ML models using PySpark and TensorFlow, boosting marketing insight accuracy and retention predictions.
- Delivered quality tools and weekly insights through dashboards and A/B test analyses using data visualization tools like Pandas, Seaborn and Matplotlib; presented findings to leadership and iterated based on 2-3x weekly feedback loops.

Research Intern **January 2020-April 2020**
Jaypee University Of Engineering and Technology [1] Guna, India

- Co-authored a research paper on transient dynamics of fibre reinforced composite plates.
- Utilized ABAQUS for finite element analysis to enhance simulation accuracy.

PROJECTS

SEC Filings NLP Pipeline |Python, NumPy, AWS, PostgreSQL, HuggingFace **April 2025-Present**

- Automated ingestion of 100K+ SEC 8-K/10-K filings into AWS S3/PostgreSQL by engineering a multithreaded Python scraper, turning a week-long manual download process into an hourly ETL job.
- Generated a 25 M-token corpus by tokenizing, entity-tagging, and normalizing filings with NLTK & spaCy, performing data cleaning and validation for large-language-model workflows.
- Fine-tuned FinBERT for tri-class risk-sentiment analysis, reaching ≈ 88 % macro-F1 via mixed-precision training on AWS GPU instances and HuggingFace Transformers, now powering downstream analytics dashboards.

FlixFriends | React.js, .NET Core, SQL Server, RESTful API, Microsoft Azure **October 2024**

- Developed a scalable social platform for discovering, rating, and reviewing TV shows and movies, featuring user profiles and real-time friend discussions.
- Utilized ASP.NET Core MVC and SQL Server for backend services and React.js for a dynamic frontend UI.
- Integrated unit testing and exception handling to improve code reliability and reduce runtime errors.

Certifications & Simulations

JP Morgan Software Engineering Simulation **April 2025**

- Solved coding challenges, participated in a mock agile workflow, applied industry-standard development practices to deliver robust, scalable solutions, and rapid prototyping under enterprise-level expectations.

IBM Certifications: Exploratory **Data Analysis** for Machine Learning; **Supervised Machine Learning: Regression**

TECHNICAL SKILLS

Languages & ML Libraries: Python, Java, SQL, R | PyTorch, TensorFlow/Keras, Hugging Face transformers, Scikit-Learn
Data Engineering & Databases: Apache Spark/PySpark, Kafka, Hadoop, BigQuery | PostgreSQL, DynamoDB, MongoDB
Cloud & MLOps: AWS (SageMaker, S3, Lambda, EC2), GCP, Docker, Kubernetes, Terraform, Git
Frameworks | React, Next.js, Django, Spring Boot, Flask, Angular, jQuery
Data Visualization & BI/Workflow: Matplotlib, Seaborn, Plotly, Tableau, Power BI, Excel | Jupyter, VS Code, Linux/macOS