Yuganshu Jain

+1 (934)255-9075 yuganshu.jain@stonybrook.edu yuganshu-jain-6047b6165/ yuganshu007 Portfolio

Education

Stony Brook University **Aug 2024 – May 2026**
Master's in Computer Science *Head of Graduate Teaching Assistant for Logic in CS*
Coursework: Distributed Systems, Systems Security, Logic in CS, Data Science, Data Visualization, HCI

Jamia Hamdard, South Delhi, India **July 2017 – Aug 2021**
Bachelor of Technology in Computer Science and Engineering *Final Year: **9.6**/10.0 CGPA*
Coursework: Distributed Systems, Advanced Computer Architecture, Advanced Database Management System, Operating System Lab, Compiler Design Lab, Advanced Java Programming, Web Technology, Data Warehousing & Mining , Big Data

Experience

Amazon **May 2025 – Aug 2025**
Software Development Engineer Intern *New York City, United States*
- Deployed a scaleable end to end ETL pipeline with an internal Java orchestrator on Elastic Map Reduce/Apache Spark clusters to run periodic jobs; processed 1,000+ hours/day of call transcripts with retries boosting performance by 2%.
- Designed Voice of Advertiser (VOA) **AI Chatbot** using Bedrock; reduced review time from 45 minutes to 2 per stakeholders call analysis using distributed joins; enabled downstream analytics with **P95 query latency < 2s** on curated datasets.
- Leveraged distributed S3 insights, **Spark RDD/DataFrame** pipelines with **Lambda, Athena, and Glue** to make AI powered Streamlit Dashboard serving 500+ Amazon Ads stakeholders enabling 10x faster data-driven decisions
- Received recognition from Amazon Ads Director for dashboard that automated early budgeting degradation alerts using Python/Plotly and data normalization, boosting campaign response time by 75 %, preventing revenue and customers loss.

RSTech Softwares **Nov 2021 – Aug 2024**
Software Engineer *Noida, India*
- Optimized full-stack system architecture using **React.js, TypeScript, PostgreSQL, Next.js, Docker**, achieving **52% reduction** in page load times enhanced user experience by mitigating latency using caching & splitting techniques
- Architected highly scalable distributed microservices system using **REST APIs** and **Apache Cassandra**, improving system performance to handle **2,500+ concurrent users** and reducing database query response times by **40%**
- Devised **Amazon SES** email-reporting pipeline with retries, resolving **344 customer issues**, boosting SLA adherence.
- Overhauled scalable B2B e-commerce platform integrating payment gateways with **QA tested** pipelines & RESTful API orchestration, reducing transaction errors by 30%, accelerating release cadence, and strengthening data integrity.
- Led cross-functional engineering team of **5 developers**, revitalized advanced **data security** measures, mentoring junior developers on best practices, and conducting code reviews, improving overall system reliability by **25%**

Infinity Haul **April 2021 – Nov 2021**
Software Development Engineer *Delhi, India*
- Applied optimization algorithms in **Java** and **Kotlin** to refactor Android modules, reducing UI stalls and improving responsiveness by **40%**, cutting CPU usage under peak loads and boosting user engagement metrics.
- Integrated Google Maps SDK with Firebase for live tracking; real-time route recomputation lowered delivery times by 20%.
- Streamlined SQLite storage and sync strategies to optimize database queries, leading to **40% reduction** in offline app load

Technical Skills

Programming Languages: Java, Python, kotlin, C++, C, , JavaScript, TypeScript, SQL, PL/SQL
Frameworks & Libraries: S3, EMR, Glue, Athena, Bedrock, SageMaker, Lambda, CloudFormation, CloudWatch, IAM, QuickSight, Streamlit, Spring Boot, Spring MVC, Flask, React, Angular, Node, RESTful APIs, GraphQL, Spark, Hadoop, Firebase, Docker, Kubernetes, Git, Maven, Jenkins,JIRA, TensorFlow, scikit-learn, PyTorch, Pandas, NumPy, Plotly, OpenCV

Research & Projects

Reddit Humor Detection & Funniness Prediction **2025**
Python, NLP, scikit-learn, TensorFlow, HuggingFace, Random Forest, LDA *Prof. Steven Skiena*
- Developed end-to-end humor detection system processing over 1M+ Reddit jokes leveraging text cleaning, TF-IDF vectorization, BERT embeddings, and Random Forest classification, resulting in 86% accuracy and dashboards and reports.
- Orchestrated clustering and regression pipeline using LDA topic modeling, TF-IDF and embeddings with MiniBatch KMeans and Ridge Regression to classify humor patterns and predict joke funniness scores achieving 0.87 R^2.

Phishing Detection System Using Advanced Machine Learning **2024**
Python, scikit-learn, XGBoost, Random Forest, Neural Networks *Link*
- Implemented intelligent threat detection system to combat rising phishing attacks, training ensemble models on 120K+ URL dataset, improving detection accuracy by **21%** while reducing false positives by **18%**.
- Built production-ready microservices architecture using FastAPI and Redis caching on Kubernetes to handle enterprise-scale security demands, supporting **1,800+ requests/sec** with **80ms** latency and comprehensive monitoring dashboards.