

Rishitosh Kumar Singh

602-642-6645 | rishitoshs@gmail.com | linkedin.com/in/rishitoshsingh | github.com/rishitoshsingh

EDUCATION

Arizona State University Master of Science in Computer Science Relevant Courses: NLP, Statistical ML, Cloud Computing, Algorithms, Image Analytics, Data Mining, Web Mining	Anticipated December 2025 C.G.P.A: 3.85 /4.00
Dr. A.P.J. Abdul Kalam Technical University Bachelor of Technology in Computer Science Relevant Courses: Data Structures, Algorithms, Operating Systems, Database Management Systems, Computer Networks	September 2020 C.G.P.A: 8.27 /10.00

SKILLS

Machine Learning & AI: Python, PyTorch, TensorFlow, scikit-learn, LLMs (RAG, MCP, fine-tuning, agentic AI), NLP
Data Engineering & MLOps: Spark, PySpark, Airflow, Kafka, MLflow, SageMaker, Docker, Kubernetes, CI/CD, automated data pipelines, model deployment
Cloud & Infrastructure: AWS (S3, EC2, Lambda, SageMaker), Snowflake, Linux, Bash scripting
Software Development: Python, C++, SQL, JavaScript, REST APIs (Flask, FastAPI), Git, system design
Visualization & Analytics: Pandas, NumPy, Matplotlib, Streamlit

EXPERIENCE

Research Assistant <i>Arizona State University</i> <ul style="list-style-type: none">Automated large-scale simulation pipelines (10K+ trials) to generate structured gameplay data for evaluating and benchmarking rule-based and other optimization strategies.Developed text analytics workflows for interview and survey data using Python, scikit-learn, and NLP methods (topic modeling, clustering) to extract actionable insights.Designed and maintained data preprocessing pipelines in Snowflake for downstream model training and analytics tasks.	February 2025 – Present Tempe, AZ
Tiger Analytics India Private Limited <i>Machine Learning Engineer</i> <ul style="list-style-type: none">Automated ML pipelines in AWS SageMaker for a UK based financial institution, cutting retraining cycles by 35%.Developed drift detection and retraining workflows using MLflow and Python, improving model reliabilityOptimized PySpark ETL pipelines with StreamSets processing millions of records daily, reducing batch runtime by 25%.Automated model monitoring and CI/CD tasks, eliminating 60% of manual effort and enhancing financial system reliability.	Hyderabad, TS June 2022 – December 2023
Senior Product Analyst (Applied ML) <i>Techlearn (formerly EduGrad) – Ashoka Systems India Pvt. Ltd.</i> <ul style="list-style-type: none">Built NLP models to analyze user behavior and recommend products, reducing customer churn from 10% to 2%.Developed and deployed EduFace, a face recognition pipeline using GStreamer, OpenCV, and Docker; reduced inference costs by 62% by running on optimized local hardware.Applied computer vision techniques to segment mentor recordings by topic, automating content evaluation and reducing manual review time.Created a Flask-based internal platform for efficient labeling and processing of user and product data.	July 2020 – June 2022 Hyderabad, TS

PROJECTS

Multi-Agent RAG and Knowledge Graph Systems <ul style="list-style-type: none">Developed a personal RAG application that indexed and summarized GitHub repositories using LangChain, vector databases, and OpenAI APIs; enabled recruiter-style Q&A about projects via a Streamlit interface.Built a semantic web mining pipeline using LinkedIn job postings and 30K resumes; extracted skills, roles, and locations to construct a knowledge graph linking candidates and jobs through skill relations.Integrated both systems into a unified retrieval framework combining LLM reasoning with graph-based entity linking for context-aware query responses.	Oct 2025
SunTransit: Transit Performance Dashboard <ul style="list-style-type: none">Designed and implemented end-to-end data pipelines using Spark, Kafka, and Python to process real-time GTFS and GTFS-RT feeds for public-sector transit operations.Containerized and orchestrated components with Docker and Airflow for automated data ingestion, scheduling, and monitoring; deployed Streamlit dashboards for real-time visualization and insights.Published an open, analysis-ready dataset of transit delay patterns, supporting reproducible research and urban planning analytics on public transport performance.	Jul 2025

CERTIFICATIONS

AWS Certified Cloud Practitioner (S73F8JT24NEE16CS)