

AKSHAT MISHRA

791 McDonald Avenue, Brooklyn, NY — am15111@nyu.edu — +1 347-864-4425 — LinkedIn: Akkey01 — GitHub: Akkey01

EDUCATION

New York University, New York

May 2026

Master's in Computer Engineering

Coursework: Machine Learning, Deep Learning, Big Data, MLops, Real Time Embedded System.

APJ Abdul Kalam Technical University

Dec 2020 - Jul 2024

Bachelor's in Computer Science

Coursework: Data Structures, Computer Networks, Operating System, Artificial Intelligence, Database Management

PROFESSIONAL EXPERIENCE

Career Launcher, Prayagraj — Data Analyst Intern

Jul 2023 - Dec 2023

- Automated ETL processes using **Python** and **SQL**, improving data accuracy by **30%** and reducing manual processing time.
- Consolidated student performance data from sources into a centralized **SQL**-based database, enhancing structured reporting and analytical insights.
- Led in developing instructional materials & dashboards using **PowerPoint**, **Tableau**, and **Excel** for training sessions, enhancing decision-making for **200+ educators**.
- Redesigned training workshops based on participant feedback, leading to a **15%** boost in post-training performance metrics.

Bringup Education, Soanbhadra — Intern ML Trainee

Feb 2023 - Apr 2023

- Created a resume screening tool using **NLTK** and **Matplotlib**, increasing recruitment efficiency by **50%** and integrating with a web application interface.
- Engineered parsing algorithms to analyze resumes on basis of **20+ parameters**, identifying top **5%** candidates and streamlining the recruitment process based on historical data which improved quality of hire metrics by **20%**.
- Optimized the data preprocessing pipeline to streamline feature extraction and formatting, resulting in a **35%** improvement.

PROJECTS

Intelligent Multimedia Processing (IMP) for Enterprises

- Orchestrated an **LLM-driven pipeline** using **RAG** architecture to enable natural language querying over multimedia transcripts (Text, Audio, Video), enhancing user interaction by **40%**.
- Curated and indexed audio-visual content using **FAISS** and **HuggingFace Longform Transformers**, optimizing retrieval accuracy with a **Top-k Precision of 92%**.
- Integrated **MLOps tools MLflow, and FastAPI** for scalable deployment, reducing model latency by **30%** in query responses.
- Collaborated with cross-disciplinary engineers to define multimedia metadata schema and annotation workflows, improving data consistency and accelerating downstream model training.
- Implemented automated **Argo CI/CD** pipelines with **Docker Compose** and **GitHub Actions**, reducing deployment time by **70%** and ensuring reproducible end-to-end ETL runs.

MTA Subway Ridership Analysis

- Formulated a scalable **Spark**-based ingestion workflow to aggregate and cleanse **30 GB** of subway ridership data, streamlining database maintenance and ensuring consistent records.
- Evaluated ridership metrics using **time-series aggregations** and **calendar-heatmap visualizations**, revealing usage trends to inform operational planning.
- Programmed **ETL** pipelines and **PySpark** jobs to automate cleaning-validation, reducing manual processing time by **30%**.
- Integrated station metadata with ridership counts into unified **dashboards**, enabling data-driven resource-allocation decisions.

Detection of GAN Generated Images

- Analyzed image processing workflows and machine learning pipelines, resulting in a **15%** improvement in classification accuracy which distinguished **CycleGAN**-generated images from real images with a **96%** accuracy (**TensorFlow, PyTorch**).
- Designed a data augmentation strategy that reduced overfitting by **25%**, enhancing the model's generalization capabilities.
- Constructed model performance through hyperparameter tuning and transfer learning, boosting precision by **20%**.

SKILLS

- Languages:** Python, R, C/C++, Java, SQL, HTML, JavaScript, Assembly, MATLAB, Bash.
- Frameworks:** NumPy, Pandas, Scikit-Learn, OpenCV, BeautifulSoup, NLTK, Matplotlib, TensorFlow, Apache, Keras, Excel, Spark, Seaborn, FastAPI, HuggingFace, FAISS, Plotly, Streamlit.
- Tools & Platforms:** Tableau, Hadoop, AWS, Power BI, Jira, Confluence, Hive, Docker, Kubernetes, MLflow, CI/CD, Git, REST APIs, Airflow, Kafka, Postman, MongoDB, MySQL, Google Cloud Platform (GCP), Azure.