Sitanshu Kushwaha Data Engineer

sak9813@nyu.edu +1 (929) 643-4480 sitanshu.tech New York, NY

linkedin.com/in/sitanshukushwaha/ github.com/Sitanshuk

EDUCATION

New York University, MS in Computer Science Sep 2023 – May 2025 | New York
Relevant Courses: Big Data, Data Science, Data Management and strategy, Cloud Computing and Big Data

University of Mumbai, BE in Computer Engineering Aug 2016 – Nov 2020 | Mumbai
9.32/10 CGPA (Ranked in **Top 10**)

PROFESSIONAL EXPERIENCE

Data Engineering Intern, NBCUniversal Jun 2024 – present | New York
- **Centralized** data storage in Unity Catalog on **Databricks**, resolving data constraints and expediting the availability of essential datasets; this redesign improved forecasting accuracy, **reduced reporting time by two days**, and enhanced the **scalability** of forecasting **workflows**.
- Spearheaded the creation of a self-service data application utilizing JavaScript, Python, and SQL; **automated** data updates that **eliminated 20 hours** of **manual** intervention per cycle, **reducing** communication **overhead** and **enhancing accuracy**, **scalability**, and **reliability**.
- **Awarded first place** in the **Innovation** Business Case **Project** for developing a **data-driven** prototype for new content formats on the Peacock platform, projected to increase audience **engagement** by **20%.**

Data Engineer Intern, NYU IT Oct 2023 – May 2024 | New York
- Implemented 10+ **ELT** pipeline with **Snowflake** & **AWS**, integrating diverse data sources (Data Warehouses & Data Marts) for enhanced analysis & decision-making.

Data Engineer, LTIMindtree Jan 2021 – Jun 2023 | Mumbai
Technical Lead, Visioncare MFF Data Engineering team - Johnson and Johnson
- Optimized **Databricks Spark** code, achieving a **30% reduction in execution time** for 50% of transformation jobs, enhancing data timeliness and **scalability** for **multi-TB datasets**.
- Implemented **event-based triggers** in **Azure Data Factory** for ETL pipelines, enhancing efficiency in handling **Big Data** from diverse sources and reducing **cloud costs by 25%**.
- Designed a **Monitoring Dashboard** in **Tableau** for **real-time data flow architecture**, enabling early identification of bottlenecks and reducing system outages by **40%**.
- Streamlined deployment by implementing **CI/CD** using **Git** and **ARM templates** (Infrastructure as Code) in **Azure**, reducing deployment time by **70%** and ensuring rapid, automated updates to **cloud infrastructure**.

SKILLS

Big Data — PySpark, Kafka, Databricks, BigQuery, Snowflake, AWS, GCP, Data Lake, ETL, NoSQL, Airflow, **Machine Learning** — Scikit Learn, Tensorflow, NLP, Neural Networks, Deep Learning, **Data Analytics** — SQL, Pandas, Numpy, Matplotlib, Seaborn, Web Scraping, Tableau, **Languages** — Python, JAVA, R., **Tools** — Git, Docker, Databricks MLOps, Retool

PROJECTS

DineSync - Real-Time Culinary Exploration in NYC, (Big Data, Spark, Kafka, MongoDB, Django)
- Engineered DineSync, a **real-time restaurant recommendation** system leveraging **Kafka** for **processing live** user check-ins, ensuring accurate seat availability data with 95% accuracy.
- Engineered a solution that automatically recommended alternative restaurants when primary choices were fully booked, resulting in a **20% decrease in drop-off rates** during peak reservation times.

Beyond the 9 to 5: A Multivariate Analysis of Work-Life Balance, Commuting, and Job Attrition,
Data Analysis, Statistical Modeling, Data Visualization
- Applied **Bayesian statistical modeling** to analyze the impact of work-life balance and commuting on job attrition, using **PyMC**, NumPy, and ArviZ.
- Visualized relationships with Matplotlib and Seaborn reveal that better work-life balance and shorter commutes correlate with lower attrition rates, mediated by job satisfaction.

Subjective Answer Evaluation using Machine Learning, (NLP, Django, TensorFlow)
- Incorporated state-of-the-art **NLP** techniques, including **BERT**, **USE**, and **Word2Vec** language models, to assess students' subjective answers by measuring **semantic similarity** against the teacher's answer.
- Published a **research paper** in International Journal for Scientific Research and Development .