

Sitanshu Kushwaha

✉ sak9813@nyu.edu 📍 New York 🔗 [linkedin.com/in/sitanshukushwaha/](https://www.linkedin.com/in/sitanshukushwaha/) 🐙 github.com/Sitanshuk

📁 PROFESSIONAL EXPERIENCE

Data Engineering Intern, NBCUniversal Jun 2024 – Present | New York

- Developed and deployed an ML-powered reporting system on Databricks, leveraging MLflow, CI/CD pipelines, and automated testing to determine the optimal cutoff time for live sports broadcasts, maximizing viewer engagement and ratings for Nielsen reporting.

AI & ML Engineer Intern, NYU IT Oct 2023 – Present | New York

- Orchestrated an ETL pipeline using Airflow and dbt to collect, transform, and store metadata in Snowflake, enabling a RAG system for LLM-powered impact analysis, allowing developers to assess schema changes, track dependencies, and retrieve insights on jobs, tables, and stored procedures.
- Developed a GenAI-powered chatbot in Streamlit with conversational memory, enabling developers to query Snowflake metadata using natural language, receive contextual follow-ups, and analyze schema modifications, leveraging LLMs and Cortex AI for intelligent retrieval and automated impact assessment.

Data Engineer, LTIMindtree Jan 2021 – Jun 2023 | Mumbai

- Led the Data Engineering team to optimize Databricks Spark code, reducing execution time by 30% for transformation jobs, improving scalability and data timeliness for multi-TB datasets.
- Implemented a CI/CD pipeline in Azure DevOps, automating build validation, testing, and deployments across environments, reducing manual effort by 70% and ensuring code reliability.

Machine Learning Engineer, Oniria Creations Mar 2020 – Dec 2020 | Remote

- Trained a text classification model using BERT embeddings as input features and TensorFlow, achieving 92% accuracy in identifying pet service provider websites from Bing search results.
- Developed an automated NLP pipeline leveraging Named Entity Recognition (NER) and a fine-tuned BERT model to extract and categorize structured data (e.g., business details, services, pricing) from pet service provider websites, building a comprehensive, queryable database.

Intern Machine learning Engineer, Toflo Fintech Consulting Dec 2019 – Mar 2020 | Mumbai

- Engineered a hybrid Recommender System for an e-commerce platform, combining collaborative and content-based filtering algorithms. Implemented A/B testing to evaluate performance.

🎓 EDUCATION

New York University, MS in Computer Science Sep 2023 – May 2025 | New York

University of Mumbai, BE in Computer Engineering Aug 2016 – Nov 2020 | Mumbai

🧠 SKILLS

Machine Learning (— Scikit Learn, Tensorflow, NLP, Neural Networks, Deep Learning, Models - Regression models, Decision Trees, XGBoost, SVM, K-means, Hierarchical Clustering, PCA, CNN, RNN, LSTMS, ARIMA), **Big Data** (Apache Spark, Kafka, Databricks, Azure, AWS, GCP, Data Lake, ETL, NoSQL), **Data Visualization** (Tableau, Matplotlib, Seaborn, Plotly), **Languages** (— Python, JAVA, SQL), **Tools** (— Git, Docker)

📁 PROJECTS

Beyond the 9 to 5: A Multivariate Analysis of Work-Life Balance, Commuting, and Job Attrition,

Data Analysis, Statistical Modeling, Data Visualization ☑

- Analyzed the influence of work-life balance and commuting distance on job attrition, with findings indicating that improved work-life balance and shorter commutes are associated with lower attrition rates, mediated by job satisfaction.

Talk2Doc - (GCP, RAG, LLM, APIs) ☑

- Architected a centralized ecosystem for students using Google Cloud Platform (GCP), integrating Retrieval-Augmented Generation (RAG) for personalized note searching and automated job application tracking.
- Built modular, event-driven highly scalable system utilizing serverless functions, queues, LLM capable of handling millions of users.