



Sitanshu Kushwaha

✉ sak9813@nyu.edu  linkedin.com/in/sitanshukushwaha/  github.com/Sitanshuk

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

PROFESSIONAL EXPERIENCE

Data Engineering Intern, NBCUniversal Jun 2024 – May 2025 | New York

- Implemented a year-round aggregation strategy in **BigQuery** for Peacock's annual user insights, processing **petabyte-scale** data incrementally to distribute computational load, reducing end-of-year query times from **hours to minutes** for **35M+** subscribers, and optimizing resource utilization and costs.
- Reengineered legacy processes by centralizing data in **Databricks' Unity Catalog** and introducing a **self-service tool**, enhancing **data governance** and improving **stakeholder transparency**. This streamlined **forecast extrapolation** across demographics, reducing communication overhead.
- Optimized** data workflows, accelerating **report delivery time** by **2 days** and eliminating **20 hours** of **manual intervention** per quarter. This improved the **efficiency** and **accuracy** of forecasting processes.

Data Science Intern, NYU IT Oct 2023 – May 2024 | New York

- Developed a decision tree model** to classify students into performance categories, achieving **85% accuracy** with **20+ features**, including attendance, grades, and socioeconomic factors.
- Conducted feature importance analysis** to pinpoint key predictors of student performance, empowering academic advisors to implement targeted support strategies to improve graduation rates.

Data Engineer, LTIMindtree Jan 2021 – Jun 2023 | Mumbai

- Optimized** Databricks Spark code, achieving a **30% reduction** in execution time for **50% of transformation jobs**, enhancing **data availability** for analysis.

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.
- Build an automated NLP pipeline using **Named Entity Recognition** and **fine-tuned BERT** models to extract and classify structured data (e.g., business details, service offerings, pricing) from pet service provider websites, creating a comprehensive 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.

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.

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 ☑ .

AWARDS

1st Place - Innovation Business Case Project, NBCUniversal Aug 2024

2nd Place - Build with AI, Google Developer Group NYC Apr 2024